STRUCTURAL AND BIOCHEMICAL STUDIES ON TAILORING ENZYMES INVOLVED IN RIBOSOMAL PEPTIDES BIOSYNTHESIS

BY

YUE HAO

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biochemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Professor Satish K. Nair, Chair, Director of Research
Assistant Professor Hong Jin
Professor Susan A. Martinis
Professor Huimin Zhao

Abstract

Human history is also a history of the ceaseless fight against various diseases, especially those caused by pathogenic infections. Both scientists and pharmaceutical companies have been dedicated in developing antimicrobial compounds to combat the micro-enemies. Despite the efforts, the resistant strains keep emerging with the introduction of new therapeutic compounds, which implies an urgent need for compounds with modified or novel scaffolds to overcome the resistance. One good source of antimicrobial compound discovery is natural products. And research on their biosynthesis is of great importance in understanding and also utilizing nature's tools to create bioactive compounds and variants.

One major class of bioactive natural products is ribosomally synthesized and post-translationally modified peptides (RiPPs). Precursor peptides and tailoring enzymes for post-translational modifications (PTMs) are encoded in the same gene cluster, which is efficient in producing peptide variants providing different precursor sequences. The peptides are first synthesized on ribosomes and then undergo extensive modifications by the tailoring enzymes to generate active, rigid and stable products. Elucidation of the biosynthetic process, especially the enzymatic mechanisms in PTMs provide the guidelines of engineering either the peptides or the tailoring enzymes to produce more effective and stable RiPP variants or derivatives.

Lanthipeptides are RiPPs bearing characteristic lanthionine (Lan) and/or methyllanthionine (MeLan) bridges. Many of them display potent antimicrobial activities, such as the class I lanthipeptides nisin and microbisporicin. However, the mechanism of one key step in their biosynthesis – dehydration – was still mysterious, until recently researchers in Wilfred van der Donk group found class I lanthipeptide dehydratase catalyzes the reaction in a tRNA-dependent

manner. To understand the structural basis of dehydration reaction, I determined the crystal structures of two class I lantibiotic dehydratases involved in the biosynthesis of nisin and microbisporicin, respectively. The structures supported the proposed two-step mechanism of dehydration reaction, illustrated the decisive factors for peptide substrate recognition, and provided clues on the tRNA-binding site based on a docking model.

Plantazolicin is a linear azol(ine)-containing peptide (LAP) that specifically inhibits the growth of anthrax causing pathogen *Bacillus anthracis*. It contains two patches of thiazole and (methyl)oxazole/methloxazoline rings that endow its rigidity. It has an N^{α}, N^{α}-dimethyl-arginine at the N-terminus, which is critical for the antibacterial activity. The *N*-methyltransferase responsible for this modification in the gene cluster is very selective to thiazole/(methyl)oxazole substrates with an N-terminal arginine. By studying its crystal structures in complex with truncated substrate analogs as well as steady-state kinetics, I demonstrated the determinants of the substrate specificity and key residues involved in interacting with substrate.

Cyanobactins are a group of small cyclic RiPPs produced by symbiotic or asymbiotic cyanobacteria. They contain rigid thiazoles and/or (methyl)oxazoles similar to LAP and additional head-to-tail cyclization on backbone to decrease its flexibility and susceptibility to proteases. Many cyanobactins are subsequently decorated with isoprenoid groups on Tyr, Ser or Thr residues for improved hydrophobicity. To understand the substrate requirements for cyanobactin prenyltransferases, I performed structural study of selected cyanobactin prenyltransferases with isoprenoid donor and various acceptors to illustrate the substrate selectivity. Examination of the substrate spectrum further supported the proposed substrate requirements based on structural information.

To my beloved parents and grandparents

Acknowledgement

The work presented here will not be accomplished without the help of so many people. I truly appreciate the great support from everyone who accompanied me during the five years of my PhD study. First, I would like to thank my advisor, Prof. Satish Nair, who recruited me, gave me exciting projects and ideas, and inspired me to think about my future direction. The experience of being supervised by him and working in his lab would be a lifelong treasure for me. I would also like to thank all my supportive collaborators: Prof. Wilfred van der Donk group, especially Manuel Ortega; Prof. Douglas Mitchell group, especially Patricia Blair; Prof. Eric Schmidt group (University of Utah), especially Elizabeth Pierce. Without their intelligence and diligence, all my work would not be as shiny as they look now. For all my past and current lab mates, I could not thank them more for making the lab like home and for making working in lab a fun thing. I want to specially thank Zhi Li for his tremendous help when I just started my PhD and was still at a lost.

I want to express my gratitude to everyone who assisted me in either research or study here. I would like to thank all the staff in LS-CAT in Argonne National Laboratory for their assistance in X-ray diffraction data collection. I would like to thank my undergraduate assistants, especially Samuel Lee and Caitlin Ondera, for their help in both my projects and lab jobs. I would like to thank Jeff Goldberg and Cara Day for their help in miscellaneous administrative affairs.

Most importantly, I want to thank my parents for their unconditional love, and for their psychological support whenever I feel intimidated or defeated. I want to thank my friends here and back in China, who have kept me company in my five years of PhD study.

Lastly, I would like to say thank you to everyone that helped me to survive my PhD.

Table of Contents

Chapter 1 Introduction 1

Chapter 2 Structural Studies on the tRNA-Dependent Lantibiotic Dehydratases 26

Chapter 3 Structural and Biochemical Studies on the *N*-Methyltransferases in Plantazolicin Biosynthesis 66

Chapter 4 Structural Studies on the Prenyltransferases in Cyanobactins Biosynthesis 101

References 134

Chapter 1 Introduction

A large part of human history is about diseases, especially the infectious diseases. 3000-year-old Egyptian mummies have had smallpox detected in them (Brachman, 2003). Ancient Greek and Egyptian documented epidemics of smallpox, leprosy and tuberculosis (Watts, 1997). It was considered that Greek physician Hippocrates first described many diseases and symptoms, and believed the dissemination of disease was through airs, waters and places (Brachman, 2003; Shulman, 2004). This belief survived and from it, the miasmas theory derived. Poisonous vapors were thought to be the origin of epidemic infectious diseases and the air contaminated by them was the vehicle of transmission (Karamanou, Panayiotakopoulos, Tsoucalas, Kousoulis, & Androutsos, 2012). This concept was held for centuries until Girolamo Fracastoro in the 1500s brought up the germ theory– infection was caused by transferable seed-like beings through three routes. However, this theory didn't become mainstream, even if the discovery of the microscope by Anton van Leeuwenhoek in 1600s inspired physicians and scientists to make connections between infections and microorganisms. In 1800s, Louis Pasteur and Robert Koch's researches eventually persuaded the majority that the microorganisms were not generated spontaneously and disease could be caused by a specific microorganism (Brachman, 2003; Karamanou, Panayiotakopoulos, Tsoucalas, Kousoulis, & Androutsos, 2012; Casanova & Abel, 2013). Since then, this remains as the dominant theory for infectious diseases.

Confronted with these pathogenic microorganisms, human beings fought back using two main approaches. The first is the development and utilization of vaccines. Vaccines resemble the pathogenic microorganisms and train the immune systems to recognize and destroy the real pathogens once they invade the body. Since the very first vaccination performed by Edward

Jenner to prevent smallpox in 1796, a number of successful vaccines have been developed and introduced to the public in the last two centuries, such as those against diphtheria, measles and polio (Stern & Markel, 2005). The second practice is to treat the pathogens with antimicrobials once the infection occurs. The concept of antimicrobial was introduced in 1900s, much later than vaccine. But the exposure to antimicrobial could be dated back to ancient Egypt, where the traces of tetracycline from dietary materials were detected in human skeletal. Another type of exposure was possibly through the natural compounds in traditional medicines (Aminov, 2010). Human have a long history of using natural resources such as plants to cure diseases. It was reported the Neanderthals in 60,000 years ago might have noticed the medicinal properties of plants (Solecki, 1975). Before the modern medicine emerged, traditional medicines from natural sources have flourished in both Occidental and Oriental countries. Nowadays, more and more natural compounds were proved to be the active components of these traditional remedies. The most famous example is the antimalarial drug artemisinin, extracted from *Artemisia* plants, which have been used as herbal remedies in China for thousands of years (Cui & Su, 2009).

The modern antibiotic era began in early 1900s. German physician and scientist Paul Ehrlich proposed and popularized the idea of "magic bullet"– an ideal therapeutic agent that selectively act on pathogens but not the host. His laboratory synthesized hundreds of organoarsenic compounds for systematic screening in syphilis-infected rabbits and discovered Salvarsan, one of the most prescribed drugs (Aminov, 2010). Later, this method also led to the discovery of Prontosil, the prototype of all sulfa drugs. This primary version of a screen program was improved into the combinatorial chemistry and high-throughput screen in 1990s, becoming one of the most important strategies in drug discovery process in pharmaceutical companies. Alexandar Fleming discovered the very first natural antimicrobial compound penicillin in 1928

and described its isolation from *Penicillium* mold and potential use (Fernandes, 2006; Aminov, 2010). The purification of penicillin opened the door to the golden era of antibiotic discovery and in the following decades a plethora of renowned antimicrobials were identified from massive screening of antimicrobial-producing microorganisms, such as the renowned vancomycin, erythromycin and teichoplanin. By 1990, around 80% of the drugs were natural products or their analogs (Li & Vederas, 2009). However, this success didn't last long because of the repeated discovery of the existing compounds. Many pharmaceutical companies then switched to the small-molecule libraries and high-throughput screening, which didn't give the expected returns, though.

Thanks to antimicrobials, millions have been cured from infectious diseases and life expectance is greatly lengthened today compared with early 1900s. But as a consequence of using antimicrobials in prescription and food industry, the new global problem now is the reemergence of infections from multi-drug resistant microorganisms. The antibiotic resistant genes such as β-lactamases were suggested present in plasmids for millions of years but the large-scale production and utilization of antibiotics exerted a strong selection pressure on them (Aminov, 2009). The horizontal transfer of these genes became the main method of acquiring antibiotic resistance for microorganisms with mutation-driven resistance happening mainly in in-host evolution. It is reported in U.S. at least 2 million people acquire infections with resistant bacteria each year and at least 23,000 people die from them (U.S. Centers for Disease Control and Prevention (CDC), 2013). Besides regulation on antimicrobial usage and prevention of infections, it is still urgent to improve old antimicrobials and discover novel antimicrobials to overcome resistance. One strategy is to generate analogs of old effective natural compounds either through semisynthesis or combinatorial biosynthesis (Li & Vederas, 2009; Bachmann,

Van Lanen, & Baltz, 2014). For instance, the large modular polyketide or nonribosomal peptide synthases provide a promising flexible assembly line for biochemical engineers to achieve increased substrate promiscuity, novel combinations and expanded product profile. A recent example is the heterologous expression of tailored erythromycin synthases to produce chiral pairs of deoxysugar analogs, of which, several demonstrated bioactivity against an erythromycin-resistant *Bacillus subtilis* strain (Zhang, Li, Fang, & Pfeifer, 2015). A second strategy is to analyze the structure of active compounds and expand the small-molecule libraries to cover more biologically relevant chemical space for high-throughput screening (Harvey, Edrada-Ebel, & Quinn, 2015). Pharmaceutical companies have designed natural-product derived or inspired chemical libraries to incorporate in the lead-like properties (Pascolutti & Quinn, 2014). And the selection of screenings containing natural products significantly increased the number of hits (Sukuru, et al., 2009). A third approach is to turn back to nature and search for new antimicrobials. One challenge is to improve the techniques of high-throughput screening for natural-product-based collections. Extractions from natural sources are complicated and the effect on the biological target is thus complex and unpredictable. For example, the effective components may be present at too low concentrations or multiple components may act synergistically. Various pre-fractionation strategies and sensitive NMR techniques have been developed to reduce the complexity and address the structure-solving problem (Harvey, Edrada-Ebel, & Quinn, 2015). On the other hand, only a small proportion of organisms in a limited variety of ecological niches have been screened for bioactive compounds, mainly the soil-dwelling *Actinomyces*. It is worth expanding to other environments such as marine species, anaerobic bacteria and many more uncultured organisms. However, the problem is, even for the species that can be cultured under natural or laboratory growth conditions, a lot of these natural

products won't be produced despite the detection of genes encoding the biosynthetic machinery. For example, in *Streptomyces spp.*, it has been shown that the antibiotic expression is regulated by a quorum-sensing system similar to the well-studied LuxI/LuxR system in *Vibrio fisheri*. A group of signaling molecules, γ-butyrolactones, are synthesized and then sensed by other cells, or by other *Streptomyces* species to induce morphological change as well as antibiotic production (Takano, 2006; Aminov, 2009). Although it is possible to explore every feasible growth condition to induce production, the discovery process would be largely impeded without exploiting the genome sequencing and the genome mining approach (Bachmann, Van Lanen, & Baltz, 2014).

With the completion of the *Streptomyces coelicolor* A3(2) sequencing project, researchers started to realize that the genome of this microbe encodes more natural-product-like biosynthetic pathways than the products previously detected (Bentley, et al., 2002). The increasing number of sequenced genomes in the database and better understanding of the genetics and enzymology of the natural products biosynthesis have significantly facilitated the identification and annotation of the biosynthetic gene clusters and the prediction of pathways and products in the last two decades (Fischbach & Walsh, 2006). Groups have been utilizing various strategies (for example, searching for typical gene structures and signature tailoring enzymes) to discover biosynthetic gene clusters from various sources, including actinobacteria (Doroghazi, et al., 2014), cyanobacteria (Leikoski, et al., 2013; Calteau, et al., 2014), anaerobic bacteria (Letzel, Pidot, & Hertweck, 2014), fungi (Khaldi, et al., 2010) and environmental DNA (Reddy, Kallifidas, Kim, Charlop-Powers, Feng, & Brady, 2012; Owen, et al., 2015). Bioinformatic tools (for example, antiSMASH (Weber, et al., 2015)), genetic engineering methods and sensitive analytical methods were also developed to speed up the identification and structural elucidation processes,

particularly for ribosomally-synthesized and post-translationally modified peptides (RiPPs), polyketides and nonribosomal peptides. The biosynthesis of both polyketides and nonribosomal are similar to fatty acid synthesis and carried out by multi-domain enzymes or enzyme complexes with small molecules as building blocks. In contrast, the biosynthesis of RiPPs starts with relatively large peptides translated on ribosomes.

RiPPs

In most cases, all the components necessary for RiPP biosynthesis are encoded in the same gene cluster, including the substrate peptide(s), all the modifying enzymes and protease for cleaving extra flanking sequences. In many instances, there are also genes for regulatory proteins and immunity proteins, which tune the expression level of the genes in the cluster and protect the producers from the potential damage caused by mature products. Transporters could also be included for exporting the mature peptides out of the cell. Under proper conditions, the gene encoding the peptide is initially transcribed and translated to precursor peptide, usually 20-110 residues in length. The precursor peptide typically contains two segments: the N-terminal leader peptide and the C-terminal core peptide. Leader peptide is required for recognition by many tailoring enzymes and for export and the core peptide is where all the post-translational modifications take place. In some classes of RiPPs, an extra recognition sequence flanks the C-terminus of the core peptide and is essential for excision and cyclization reactions (Arnison, et al., 2013). The tailoring enzymes encoded in the same gene cluster process the precursor peptide by installing various functional motifs, such as dehydrated amino acids, heterocycles, halogens, sugar moieties, and disulfide bonds. These modifications endow the substrate peptides with

structural rigidity, proteolytic stability and target-binding properties. The leader peptide is hydrolyzed afterwards and the mature peptide is secreted out of the cell (Fig. 1.1).

A structural biology approach in elucidating RiPP biosynthesis is of considerable value in the sense that structural information on the enzyme-substrate interactions not only supports or further details the reaction mechanism but also provides the rationale for engineering the biosynthetic system. RiPPs possess attractive properties as drug candidates but are generally hard for total chemical synthesis. Given that natural variants occur in RiPP biosynthesis, research on the substrate tolerance of the machinery is crucial in generating more unnatural variants for bioactivity assessment. As an essential complement to the codon reprogramming experiments of the precursor peptide, structure study on the modifying enzymes illustrates what elements are necessary for substrate recognition and efficient turnover, and how much change on either substrate or enzyme could be tolerated to achieve balance in efficient substrate processing and expected product diversity.

As mentioned above, the biosynthetic features of RiPPs provide a good platform for obtaining unnatural novel variants by designing new precursor peptides for the enzymes. Numerous studies on the RiPP biosynthetic machinery from various classes showed that the tailoring enzymes are quite tolerant on the peptide substrates and thus rational design of the peptides was attempted. A collection of RiPP variants with natural or unnatural amino acids incorporated were successfully produced through either heterologous expression of the gene cluster or simply one-pot *in vitro* enzymatic reactions, which further confirmed the potential of RiPPs as a reservoir for drug discovery (Li, Zhang, & Kelly, 2011; Pan & Link, 2011; Tianero, Donia, Young, Schultz, & Schmidt, 2012; Deane, Melby, Molohon, Susarrey, & Mitchell, 2013; Sardar, Lin, & Schmidt, 2015).

RiPPs have a wide variety of bioactivities, inclusing antimicrobial, antifungal, antiviral, anticancer, and cytotoxic. The bioactivity is believed to endow producer strains with the growth advantage by killing or inhibiting the competing microbes and this free gift from nature turns out to be a powerful tool for human beings to fight against diseases (Nolan & Walsh, 2009). The bacteriocin glycocin F secreted by *Lactobacillus plantarum* KW30 has been shown to inhibit the growth of a related *Lactobacillus* strain (Stepper, et al., 2011) and a panel of strains in the genera *Streptococcus*, *Enterococcus* and *Bacillus*. Lanthipeptides are generally active against Gram-positive bacteria. The pinensins produced by Gram-negative bacterium *Chitinophaga pinensis* are the first lanthipeptides that exhibited significant inhibitory activity against filamentous fungi and yeasts, but weak antibacterial activity (Mohr, et al., 2015). Lanthipeptide labrynthopeptin A1 potently inhibit the entry of HIV and HSV to cells and block the viral cell-cell transmission with no significantly induced inflammation response (Férir, et al., 2013). Lanthipeptide cytolysin produced by *Enterococus faecalis* species display hemolytic activity and enhance enterococcal virulence in pathogenic infections (Coburn & Gilmore, 2003). The cyclic cyanobactins also demonstrate antibacterial, antiviral and anticancer activities. Some are highly cytotoxic against human cancer cell lines and others show inhibition to the drug efflux system (Houssen & Jaspars, 2010).

Based on structural characteristics of the mature peptides and the biosynthetic machinery, RiPPs can be further grouped into different classes. They feature different types of post-translational modifications, yet each type of modifications is not solely found in one specific class (Fig. 1.2). For example, the dehydration reaction on Ser/Thr residue is not exclusively used in lanthipeptides biosynthesis to prime the formation of thioether linkages. It is also a common strategy for linaridins and thiopeptides to obtain dehydrated amino acids. The azoles in

thiopeptides are also found in linear azol(in)e-containing peptides and cyanobactins, whereas the mechanisms of heterocyclization reactions and the compositions of the biosynthetic machinery differ from one another. Another example is the generation of head-to-tail macrocycle in cyanobactins and orbitides biosynthesis through excision of the C-terminal sequence followed by macrocyclation. Although the two proteases are quite distinct from each other, they utilize the same catalytic residues and achieve the same result. Here, I would like to describe more of the structural characteristics and biosynthetic pathways on lanthipeptides, linear azol(in)e-containing peptides and cyanobactins, on which my thesis research focused.

Lanthipeptides

Lanthipeptides are ribosomal peptides containing characteristic *meso*-lanthionine (Lan) or (*2S,3S,6R*)-3-methyllanthionine (MeLan) crosslinks. Lan is composed of two alanines linked by sulfur on β carbons, forming a thioether linkage; MeLan has one more methyl group on the β carbon of one of the alanines (Fig. 1.3 A). Lan residue was first isolated in 1941 (Horn, Jones, & Ringel, 1941), but the first lanthipeptide nisin was described in 1928 (Rogers, 1928) with its structure determined in 1971 (Gross & Morell, 1971). The Lan and MeLan are introduced by two reactions: dehydration and cyclization. They could be catalyzed by one synthetase or by separate dehydratase and cyclase. The Ser and Thr residues in the core peptide are first dehydrated to form 2,3-didehydroalanine (Dha) and (*Z*)-2,3-didehydrobutyrine (Dhb), respectively. The α,β-double bond is subsequently attacked by the side chain of Cys to generate an enolate structure and protonation of the enolate produces the Lan or the MeLan. In some lanthipeptides, the enolate further attacks a Dha, resulting in the labionin (Lab) structure (Fig. 1.3 A) (Knerr & van der Donk, 2012; Arnison, et al., 2013). Lanthipeptides with antibiotic activity are termed

lantibiotics (Kellner, et al., 1988). Traditionally lantibiotics are discovered from Gram-positive producers and active against Gram-positive bacteria such as drug-resistant *Staphylococci* and *Enterococci*. The mechanism of action is believed to inhibit cell wall biosynthesis by sequestering precursor molecules and disrupt membrane integrity through pore formation. Recently, the genome mining projects have identified biosynthetic gene clusters from Gram-negative bacteria (Li, et al., 2010; Singh & Sareen, 2014) and the activity spectrum is also broadened to cytotoxicity and antifungal activity.

Lanthipeptides are categorized into four classes based on the enzymatic mechanisms of installing Lan/MeLan (Fig. 1.3 B). For class I lanthipeptides, two separate enzymes, dehydratase LanB and cyclase LanC, are responsible for catalyzing the formation of thioether crosslinks. LanB (Fig. 1.3 B) proteins usually contain around 1,000 amino acids and share no homology with any other known proteins. It was postulated previously that LanB dehydration activity involves the phosphorylation of the Ser/Thr side chains. The actual mechanism of dehydration reaction remained mysterious for many years until Garg *et al*. observed glutamylated peptide intermediates instead of phosphorylated intermediates in *in vitro* reconstitution study when using a LanB mutant deficient in elimination reaction (Garg, Salazar-Ocampo, & van der Donk, 2013). Further investigation showed LanB activates the side-chain hydroxyl groups of Ser and Thr by transferring the glutamyl group from Glu-tRNA^{Glu} to them and subsequently eliminates it together with the α proton abstraction to generate Dha and Dhb (Ortega, Hao, Zhang, Walker, & van der Donk, 2015). LanC proteins are around 400 residues in length and share 20-30% sequence identity across the family (Knerr & van der Donk, 2012). The mechanism and structure study of the LanC enzyme in nisin biosynthesis revealed that the zinc ligand coordinated by the

Cys-Cys-His triad activates the thiol group of the substrate Cys to facilitate the nucleophilic attack onto Dha and Dhb (Li, Yu, Brunzelle, Moll, van der Donk, & Nair, 2006).

For class II lanthipeptides, the dehydration and cyclization are catalyzed by a bifunctional enzyme LanM. LanM contains an N-terminal dehydration domain, with no homology with LanB, and a C-terminal LanC-like cyclization domain, with conserved zinc-binding ligands (Fig. 1.3 B). Mechanistic studies of various LanMs showed both dehydration and cyclization are leader peptide dependent whereas the activity of the each domain doesn't require the presence of the other. Dehydration is adenosine triphosphate (ATP) and magnesium (Mg^{2+}) dependent; cyclization is zinc dependent as predicted. Further dissection of the dehydration reaction revealed the phosphorylation of Ser and Thr side chains and the ADP-dependent phosphate elimination (Xie, Miller, Chatterjee, Averin, Kelleher, & van der Donk, 2004; Chatterjee, et al., 2005; Li, et al., 2010; Dong, Tang, Lukk, Yu, Nair, & van der Donk, 2015). Despite the intertwined dehydration and cyclization, the directionality of each reaction type has been proved consistent for one LanM, but could be either N-to-C or C-to-N for different LanMs (Lee, Ihnken, You, McClerren, van der Donk, & Kelleher, 2009; Mukherjee & van der Donk, 2014). The first structure of a LanM protein revealed an unexpected lipid kinase-like dehydratase domain and a predicted NisC-like cyclization domain, with peptide-binding grooves proposed in each of them. Compared with lipid kinase, the dehydratase domain connects the conserved secondary elements in a noncanonical way with the phosphorylation site and elimination site adjacent to each other.

The majority of the class III lanthipeptides contains the two-ring Lab structures besides Lan crosslinks (Meindl, et al., 2010). The polycycles are installed by a multifunctional synthetase LanKC, also in a leader peptide dependent manner (Müller, Ensle, Krawczyk, & Süssmuth, 2011; Wang & van der Donk, 2012). LabKC contains N-terminal pSer/pThr lyase domain, a Ser/Thr

kinase domain and C-terminal cyclase domain (Fig. 1.3 B). The precursor peptide is first phosphorylated on the kinase domain and then phosphate is eliminated by the lyase domain to generate dehydrated intermediate. Compared to LanMs, the kinase domain of LanKC displays relaxed phosphate donor selectivity. Some LanKCs accept all NTPs and dNTPs as cofactor while others prefer certain types (Völler, Krawczyk, Pesic, Krawczyk, Nachtigall, & Süssmuth, 2012; Wang & van der Donk, 2012; Krawczyk, Völler, Völler, Ensle, & Süssmuth, 2012). The corresponding NDPs and dNDPs are not required for the lyase domain to eliminate phosphate in contrast to LanMs. Interestingly, the cyclase domain harbors no zinc-binding residues and the cyclization reaction towards Lan or Lab works with no need for energy or any metal. The directionality of the reactions is believed C-to-N and the deficiency in the C-terminal Lab ring formation completely abolish the N-terminal ring cyclization (Krawczyk, Ensle, Müller, & Süssmuth, 2012; Wang & van der Donk, 2012).

The founding member of class IV lanthipeptides was the putative lantibiotic venezuelin produced by a cryptic gene cluster in *Streptomyces venezuelae*. The predicted synthetase LanL successfully catalyzed *in vitro* dehydration and cyclization reaction on the recombinantly expressed peptide substrate with the addition of ATP and Mg^{2+} (Goto, Li, Claesen, Shi, Bibb, & van der Donk, 2010). Protein sequence analysis of LanL showed a domain organization similar to LanKC and the three reactions – phosphorylation, elimination and cyclization – were readily assigned to each domain through biochemical studies with truncated enzymes (Goto, Okesli, & van der Donk, 2011). One obvious distinction from LanKC is that the cyclase domain in LanL retains the zinc-binding ligands conserved in LanC and LanM. Thus the cyclase activity is believed to be zinc dependent.

The classic example of lanthipeptide is nisin produced by *Lactococcus lactis*, the longest-known lanthipeptide. It serves as the model system for studying the mechanisms and structures of the biosynthetic enzymes. A recent addition to class I lanthipeptide is microbisporicin produced by *Microbispora coralina*, which displayed potent activity against drug-resistant Gram-positive bacteria comparable to nisin. In Chapter 2, I describe more about the their structure features and biosynthetic pathways. I also present the structure studies on the LanB enzymes involved in nisin and microbisporicin biosynthesis. LanB has similar molecular weight with other lanthipeptide synthetases but conducts only dehydration reaction. It is intriguing to understand how the structure elements are organized, how LanB recognizes its substrate and how it uses a less prevalent resource Glu-tRNA^{Glu} rather than ATP to achieve the two-step dehydration reaction.

Linear azol(in)e-containing peptides

The very first linear azol(in)e-containing peptide (LAP) discovered is the cell-associated hemolysin streptolysin S (SLS) described in 1938 (Todd, 1938). Although it has been confirmed as a LAP since the discovery of its biosynthetic gene cluster and preliminary functional inspection of the proposed biosynthetic enzymes, the structure of SLS is still unknown (Mitchell, et al., 2009). The extensive and elaborate investigation of LAP synthetase started in 1990s, when a comprehensive study on the installation of azole rings in microcin B17 was published in *Science* (Li, Milne, Madison, Kolter, & Walsh, 1996). Later on, a growing number of gene clusters similar to microcin B17 were identified through bioinformatic study (Lee, et al., 2008; Melby, Nard, & Mitchell, 2011) and the enzymatic mechanisms on LAP biosynthesis were

unveiled gradually through further examination of the microcin B17 system or other representative ones.

A typical LAP gene cluster encodes the precursor peptide A, the azole installing complex B, C and D, protease and transporters (Fig. 1.4 A). Some clusters also encode other modifying enzymes and auxiliary apparatus such as transcription regulator and immunity protein. At first, the precursor peptide is as well transcribed on ribosome. Then the azole rings are installed by the BCD complex, but the function assignment of each protein in the complex took over a decade. The B protein (McbC for microcin B17) was first confirmed as a flavin mononucleotide (FMN)-dependent dehydrogenase to oxidize azolines to azoles (Li, Milne, Madison, Kolter, & Walsh, 1996; Milne, et al., 1999) (Fig. 1.4 B). From primary sequence analysis and *in vitro* reconstitution of microcin B17 synthetase, C protein (McbB) was initially perceived to be catalytically active in azoline ring formation and D protein was proposed for substrate recognition. Additional attempts on either decoupling the heterocyclization and dehydrogenation or dissecting the former reaction in microcin B17 biosynthesis was in vain since omitting any of the three components resulted in a completely dead complex for precursor processing (Milne, et al., 1999). The roles of C and D protein became clear when the BalhCD system from *Bacillus* sp. Al Hakam was successfully reconstituted *in vitro* (Melby, Dunbar, Trinh, & Mitchell, 2012). Unlike the McbBCD, the heterocyclization and dehydrogenation in Balh system are readily independent of each other although the trimeric complex formed is as tight as McbBCD (Dunbar, Melby, & Mitchell, 2012; Melby, Li, & Mitchell, 2014). Further study revealed that the D protein alone is enough to catalyze azoline ring formation and ATP is consumed stoichiometrically to activate the substrate amide backbone, not for activating the enzyme or energizing the substrate displacement. The side chain hydroxyl of Ser/Thr or the thiol of Cys first

gets deprotonated and attack the carbonyl carbon of the preceding amide bond, forming a hemiorthoamide intermediate. The resulting nucleophilic oxygen of the tetrahedral structure attacks the γ phosphate of ATP to form phosphorylated intermediate. The subsequent elimination of the phosphate generates an azoline ring (Dunbar, Melby, & Mitchell, 2012; Dunbar & Mitchell, 2013) (Fig. 1.4 B). The heterocyclase D processes Cys much more efficient than Ser/Thr and also demonstrates regioselectivity (Belshaw, Roy, Kelleher, & Walsh, 1998; Melby, Dunbar, Trinh, & Mitchell, 2012; Bentley, et al., 2002). The protein C, however, is not redundant. It binds heterocyclase D, regulates ATP binding and ensures proper catalysis (Dunbar, Chekan, Cox, Burkhart, Nair, & Mitchell, 2014). The importance of leader peptide has been demonstrated through mutational studies in different LAP systems and protein C is demonstrated to provide leader-binding site (Roy, Kim, Baleja, & Walsh, 1998; Mitchell, et al., 2009; Dunbar, Chekan, Cox, Burkhart, Nair, & Mitchell, 2014). The directionality of both heterocyclization and dehydrogenation has been thoroughly studied and a general N-to-C or C-to-N direction was observed (Kelleher, Hendrickson, & Walsh, 1999; Melby, Dunbar, Trinh, & Mitchell, 2012; Melby, Li, & Mitchell, 2014). The leader is then cleaved by a protease and the core peptide either undergoes more modifications or is exported as mature product.

LAPs display a variety of bioactivities. Microcin B17 inhibits bacterial DNA gyrase (Vizán, Hernández-Chico, del Castillo, & Moreno, 1991), streptolysin S and clostridiolysin S are hemolytic (Todd, 1938; Gonzalez, et al., 2010), and plantazolicin produced by *Bacillus amyloliquefaciens* FZB42 is a narrow spectrum antibiotic specifically against closely related *Bacilli* (Scholz, et al., 2011). Plantazolicin (PZN) contains two polyazolic patches and apart from the presence of a non-oxidized methyloxazoline ring, it also has a N^{α}, N^{α}-dimethyl-arginine at N-terminus (Kalyon, et al., 2011; Molohon, et al., 2011). This dimethylation is crucial for its

bioactivity since the desmethylPZN completely loses its ability to inhibit the growth of the anthrax-causing bacterium *Bacillus anthracis* (Molohon, et al., 2011). The methyl groups are installed by a *S*-adenosylmethionine (SAM)-dependent α-*N*-methyltransferase PznL and it works selectively on azole containing substrates (Lee, et al., 2013). In Chapter 3, I present my work on the structural study of two PznL homologs and explain the structural determinants for the substrate selectivity. A minimal substrate could also be proposed from structural information. I also present the kinetic analysis on peptide substrate analogs with both the wildtype enzyme and several mutants.

Cyanobactins

Cyanobactins are small cyclic peptide produced by free-living cyanobacteria or symbiotic ones within ascidians, with several recently discovered members with linear structures (Sivonen, Leikoski, Fewer, & Jokela, 2010; Leikoski, et al., 2013). The head-to-tail macrocyle typically contains six to ten amino acids with varying numbers and combinations of azoles and azolines. Some cyanobactins have additional decorations such as prenylation, methylation and disulfide bridge. The first discovered cyanobactins were ulicyclamide and ulithiacyclamide described in 1980 (Ireland & Scheuer, 1980), and the latter showed high antitumor activity when tested with both leukemia cell lines and solid tumors (Houssen & Jaspars, 2010). The search for marine bioactive compounds in the following decades led to the discovery of more than a hundred cyanobactins. Their ribosomal origin became clear when the first cyanobactin gene cluster and biosynthetic pathway were described for patellamide in 2005 (Schmidt, et al., 2005).

The cyanobactin precursor peptides and enzymes necessary for maturation are encoded in the same gene cluster. There are more than one precursor peptide gene E in the cluster and each

precursor peptide E has the potential to produce a variety of mature cyanobactins (Donia, Ravel, & Schmidt, 2008; Donia & Schmidt, 2011). The peptide E contains a N-terminal leader peptide, recognition sequence I (RSI), recognition sequence II (RSII) and core peptide, followed by recognition sequence III (RSIII); multiple repeats of RSII, core peptide and RSIII could occur with varied core sequence (Fig. 1.5 A). Compared with the hypervariable core peptide, other elements are relatively conserved/homologous, indicating their role in guiding correct post-translational modifications (Donia, et al., 2006; Sardar, Pierce, McIntosh, & Schmidt, 2015). The precursor peptide first encounters the heterocyclase D and has thiazoline/(methyl)oxazoline rings installed in a C-to-N directionality with a mechanism similar to the D enzyme in LAP biosynthesis (McIntosh & Schmidt, 2010; McIntosh, Donia, & Schmidt, 2010; Koehnke J. , et al., 2013). The notable difference is the oxygen of hemiorthoamide tetrahedral intermediate attacks the α-phosphate of ATP instead of γ-phosphate to generate an adenylated intermediate, and then the AMP gets eliminated (Koehnke J. , et al., 2013; Koehnke, et al., 2015) (Fig. 1.5 B). Most D enzymes don't distinguish Cys and Ser/Thr with the exception of TruD, which preferentially works on Cys (McIntosh & Schmidt, 2010; McIntosh, Donia, & Schmidt, 2010). The leader peptide was postulated to be necessary for heterocyclization based on the general rule of RiPP biosynthesis. However, both biochemical and structural studies suggested it is not the leader but the conserved RSI is indispensable for substrate recognition and productive binding (Koehnke J. , et al., 2013; Sardar, Pierce, McIntosh, & Schmidt, 2015; Koehnke, et al., 2015). Its β strand structure forms extensive hydrogen-bonding interaction with the β strand of a α+β domain in the D enzyme and this domain is therefore assigned as recognition sequence-binding domain. A universal leader peptide recognition motif was then proposed based on this example together with other RiPP enzyme-peptide complex structures (Koehnke, et al., 2015; Burkhart,

Hudson, Dunbar, & Mitchell, 2015). Next, the subtilisin-like serine protease A recognizes RSII and cuts the flanking sequence N-terminal to the core (Lee, McIntosh, Hathaway, & Schmidt, 2009; Agarwal, Pierce, McIntosh, Schmidt, & Nair, 2012; Houssen, et al., 2012) (Fig. 1.5 A). A similar protease G recognizes RSIII and cuts the C-terminal flanking sequence and cyclize the core peptide (Donia, Ravel, & Schmidt, 2008) (Fig. 1.5 A). Not only the RSIII but also the restrained cyclic structure of the last residue in the core peptide is suggested to be critical for macrocyclization activity (McIntosh, Robertson, Agarwal, Nair, Bulaj, & Schmidt, 2010; Koehnke J. , et al., 2012). It is believed an acyl-enzyme intermediate is formed in the cyclase reaction as other serine proteases and the aminolysis outcompetes hydrolysis to promote cyclic product accumulation rather than its linear counterpart (Agarwal, Pierce, McIntosh, Schmidt, & Nair, 2012). To impose more restraints on the heterocycles, a FMN-dependent oxidase dehydrogenates the azoline rings to corresponding azoles, but limited research has been done on the details and the exact timing of the reaction remains unknown (Houssen, et al., 2014). In some cases, the core peptide undergoes more modifications, such as methylation by a SAM-dependent methyltransferase and prenylation by a dimethylallyl pyrophosphate (DMAPP)-dependent prenyltransferase F. To further assess the potential of this naturally occurred combinatorial library, researchers were able to design the sequences of the core peptide to build unnatural cyanobactin libraries either by heterologous expression or by *in vitro* one-pot enzymatic synthesis (Donia, et al., 2006; Donia, Ravel, & Schmidt, 2008; Tianero, Donia, Young, Schultz, & Schmidt, 2012; Houssen, et al., 2014; Ruffner, Schmidt, & Heemstra, 2015; Sardar, Lin, & Schmidt, 2015).

Based on the gene organization in the operon(s), the cyanobactin gene clusters are classified into four genotypes (Donia & Schmidt, 2011) (Fig. 1.5 C). Genotype I is the most classic one

with protease A, heterocyclase D, prenyltransferase F, and protease G with the oxidase fused at N-terminus. Genotype II doesn't contain heterocyclase and oxidase. Correspondingly, there is no RSI in precursor sequences in the precursor peptides and no heterocycles in the mature products. For proper cleavage of the C-terminal flanking sequence and macrocyclization, a proline appears as the last residue of core peptide. Genotype III contains every necessary enzyme as genotype I but with a standalone oxidase. Likewise, genotype IV also has the standalone oxidase while a methyltransferase is fused to protease G, indicating further decoration on the final products. Recent genome mining study suggested more types of gene organization such as the fusion of a methyltransferase with the protease A and the fusion of a methyltransferase with the prenyltransferase F (Leikoski, et al., 2013). Short linear cyanobactins were also discovered in this study.

The prenyltransferase F is not essential in cyanobactin biosynthesis but the prenylation of the polar groups certainly increases the lipophilic properties of cyanobactins. The TruF in *tru* pathway selectively reverse-prenylates Ser/Thr side chains (Donia, Ravel, & Schmidt, 2008). The LynF in *lyn* pathway reverse-prenylates the hydroxyl group of Tyr and the subsequent nonenzymatic Claisen rearrangement results in a forward-prenylated product (McIntosh, Donia, Nair, & Schmidt, 2011). PagF and ThcF in *pag* ant *thc* pathways forward-prenylate the hydroxyl groups of Tyr in the peptide substrate. The PirF in *pir* pathway is unusual in the sense it doesn't prenylates but geranylates the Tyr. In Chapter 4, I present structural studies on PagF and PirF and propose their substrate requirements. I also tested the substrate specificity of ThcF and PirF. The results show the cyanobactin prenyltransferases have broad substrate spectra and are potential tools to prenylate any peptide to increase peptide lipophilicity.

Figures

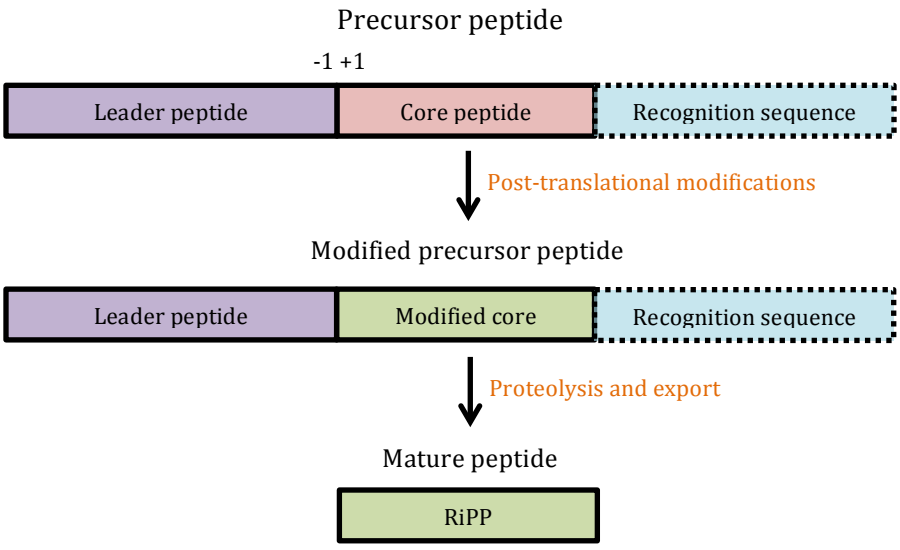

Fig. 1.1 General biosynthetic pathway of RiPPs. The ribosomally synthesized precursor peptide contains core peptide for post-translational modifications. The N-terminal leader peptide and C-terminal recognition sequence guide tailoring enzymes to perform modifications accurately. After cleavage of the flanking sequences, the modified core is exported as mature peptide.

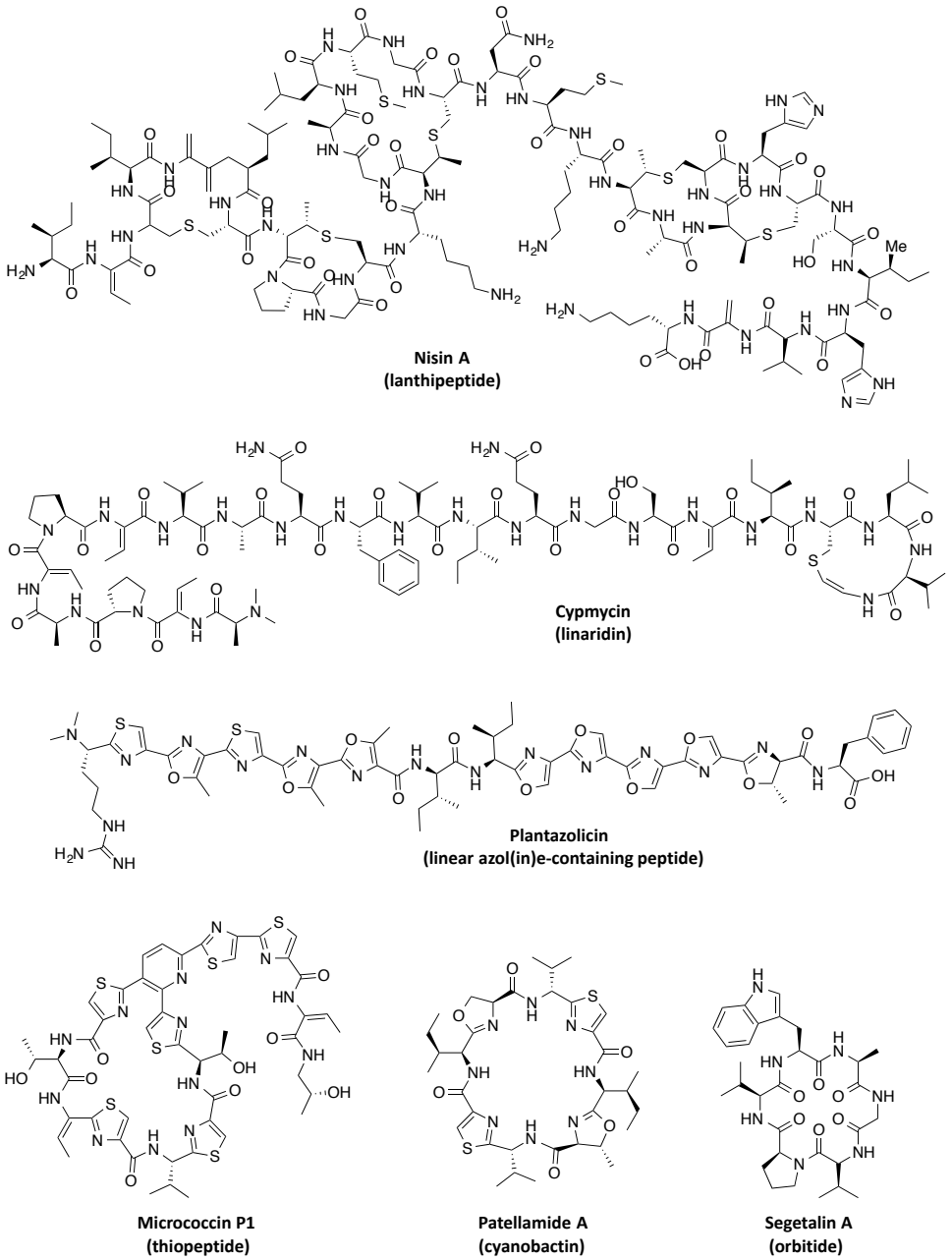

Fig. 1.2 Representative compounds from different RiPP classes. Nisin A is from lanthipeptide class; cypemycin is the first member discovered in linaridin class; plantazolicin is a linear azol(in)e-containing peptide; Micrococcin P1 is a thiopeptide; patellamide A is a representative cyanobactin; segetalin A is an orbitide.

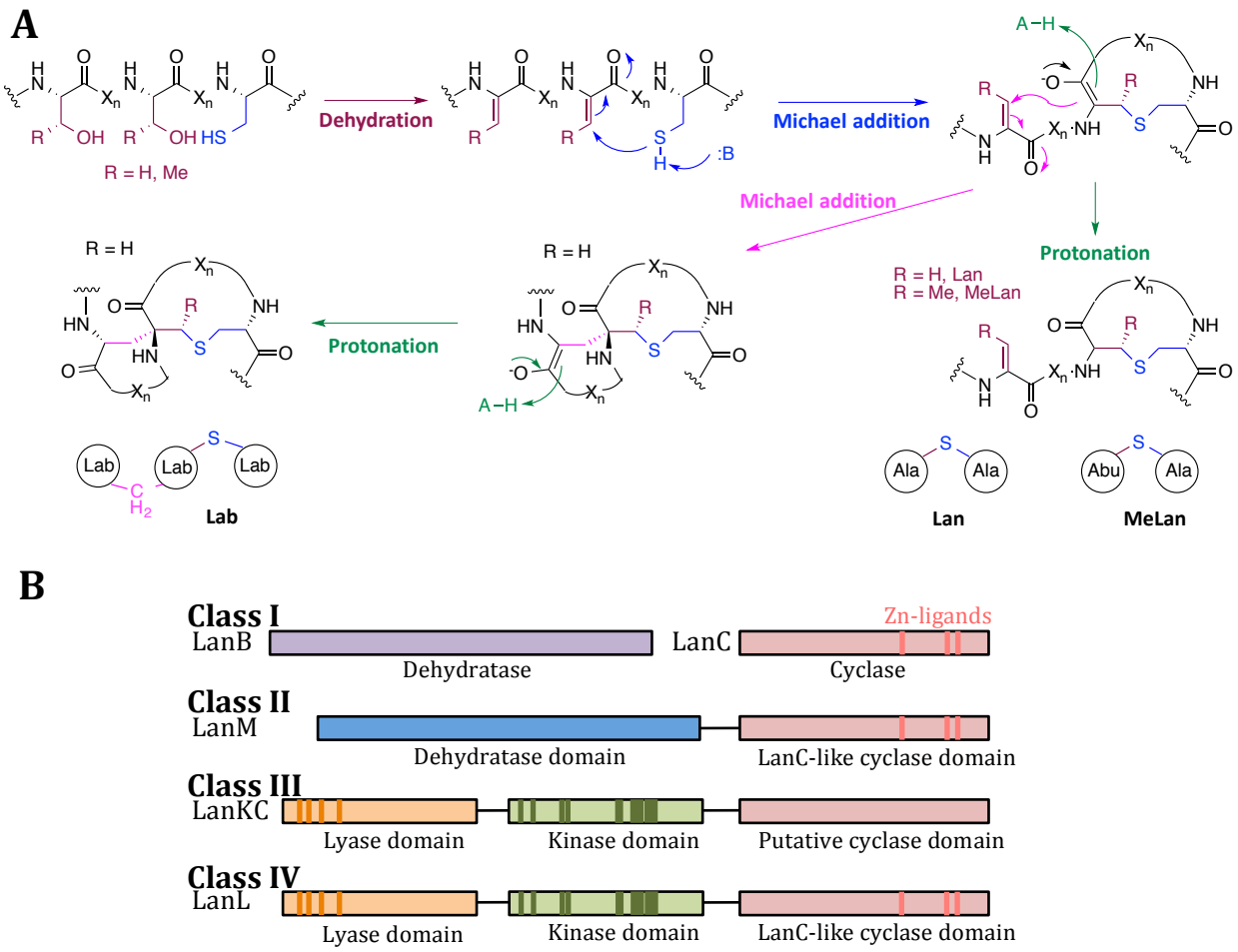

Fig. 1.3 **A**. Biosynthetic pathway to (Me)Lan and Lab formation. Ser and Thr are first dehydrated to form Dha and Dhb, respectively. Subsequent attack by the Cys side chain generates an enolate. Protonation of the enolate produces (Me)Lan. The further attack of the enolate to a Dha produces Lab. **B**. Four classes of lanthipeptides defined by the different biosynthetic machinery and the enzymatic mechanisms of installing (Me)Lan.

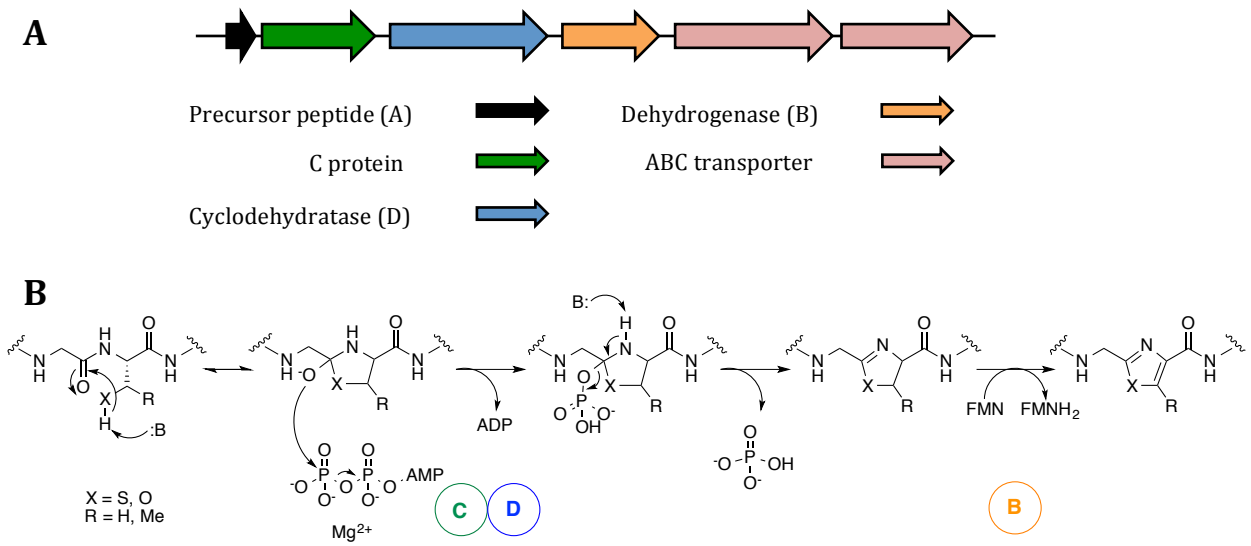

Fig. 1.4 **A.** A representative gene cluster of linear azol(in)e-containing peptide. It contains precursor peptide A, the cyclodehydratase (heterocyclase) D, a leader peptide-binding protein C, dehydrogenase B and transporters. **B.** Biosynthetic pathway to azole formation. Nucleophilic attack by the side chain of Ser/Thr/Cys onto the preceding carbonyl generates a hemiorthoamide intermediate. CD protein complex phosphorylates the intermediate to facilitate the elimination reaction rather than the cleavage of the amide bond. The resulting azoline is further oxidized by the FMN-dependent dehydrogenase B to generate azole.

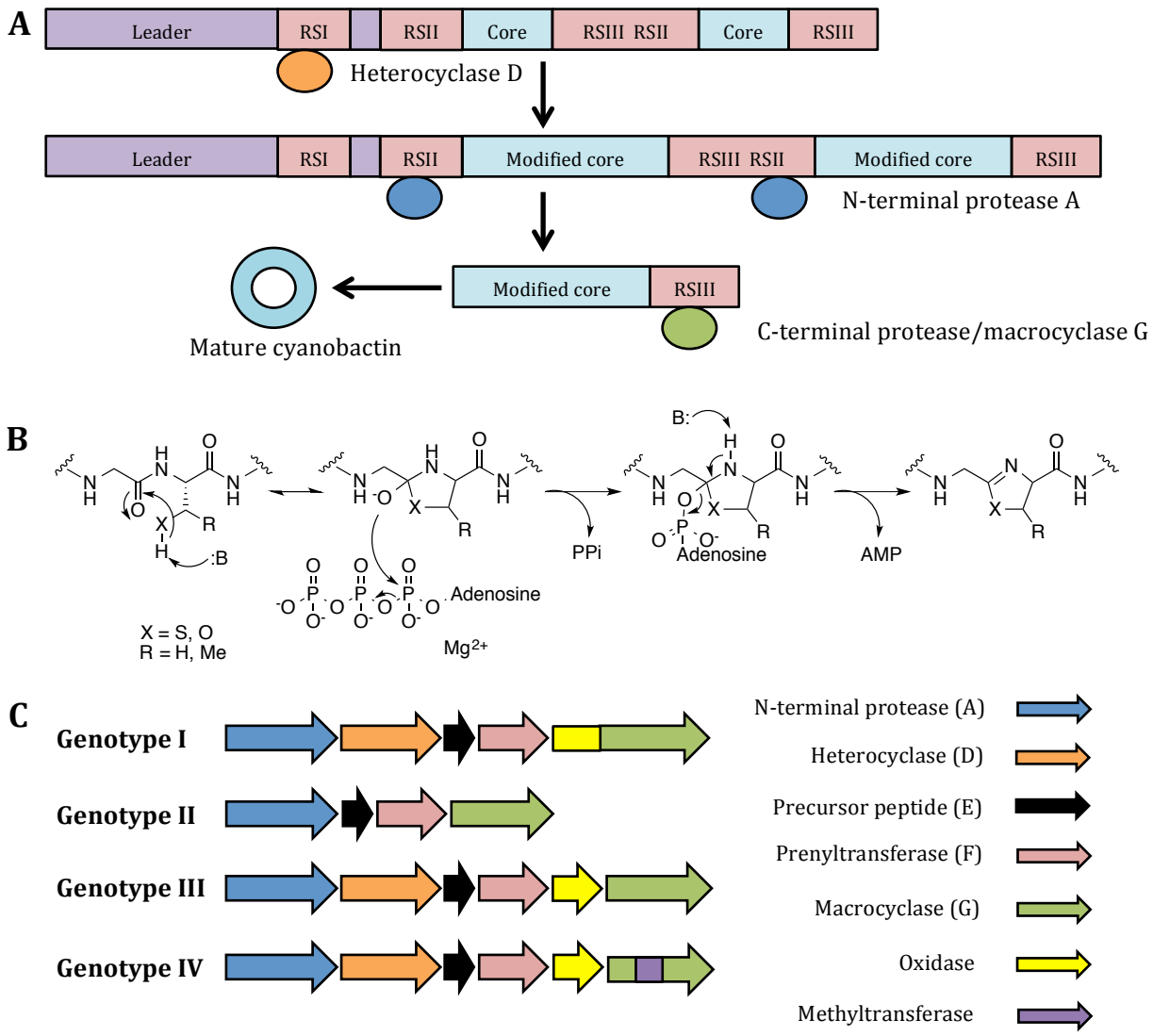

Fig. 1.5 **A.** The general biosynthetic pathway of cyanobactins. The ribosomally translated precursor peptide contains multiple core peptides and RSs for enzyme recognition. The heterocyclase D recognizes RSI and catalyzes the formation of azolines from Ser/Thr/Cys. The N-terminal protease cleaves the N-terminal flanking sequence. The macrocyclase cleaves the C-terminal flanking sequence and cyclize the modified core peptide. **B.** The biosynthetic pathway to azoline formation. The mechanism of azoline installation is similar to that of LAP biosynthesis. The side chain of Ser/Thr/Cys attacks the preceding carbonyl to generate a

hemiorthoamide intermediate. The intermediate is then adenylated instead of phosphorylated and elimination of AMP facilitates the azoline generation. **C.** General representation of four genotypes of cyanobactins.

Chapter 2 Structural Studies on the tRNA-Dependent Lantibiotic Dehydratases[1]

2.1 Structural Study of Lantibiotic Dehydratase NisB in Nisin Biosynthesis

Introduction

Nisin, produced by *Lactococcus lactis*, is the longest known lantibiotic. Its antimicrobial activity against Gram-positive bacteria was first described in 1928 (Rogers, 1928). It was partially purified and analyzed in 1944 and its structure was determined in 1971 (Mattick & Hirsch, 1944; Gross & Morell, 1971). Although nisin has the potential as a therapeutic agent in pharmaceutical industry, so far its application is mainly in food industry to combat food-borne pathogens such as *Listeria monocytogenes*. Made by a food-grade organism, nisin has been used as a safe and effective food preservative for almost 50 years without any significant resistance developed. Due to its important industrial application, nisin has been intensively studied on every aspect, including its mode of action and biosynthesis.

Nisin displays potent antibacterial activity against a wide range of Gram-positive bacteria and their spores, including multidrug-resistant *Streptococcus pneumoniae*, methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant *Enterococci* (Severina, Severin, & Tomasz, 1998). By binding to lipid II, the final intermediate to peptidoglycan biosynthesis, nisin exhibits its activity in two mechanisms: inhibiting cell wall synthesis and forming pores on cell membrane. The N-terminal A and B rings of nisin cage the pyrophosphate moiety of lipid II and block the access of transglycosylase. Thus lipid II is sequestered from cell wall synthesis at the

[1] Part of this chapter is adapted from the following published article:

Ortega, M. A., Hao, Y., Zhang, Q., Walker, M. C., van der Donk, W. A., & Nair, S. K. (2015). Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. *Nature, 517* (7535), 509-512.

cell division site. Nisin is also positioned in a transmembrane orientation by binding to lipid II and they form a hybrid pore in the membrane, disrupting membrane potential and causing efflux of metabolites and cessation of biosynthesis (Brotz, et al., 1998; Breukink, Wiedemann, van, Kuipers, Sahl, & de Kruijff, 1999; van Heusden, de Kruijff, & Breukink, 2002; Breukink, et al., 2003; Hasper, de Kruijff, & Breukink, 2004; Hsu, et al., 2004; Hasper, et al., 2006).

The identification and characterization of nisin biosynthetic gene cluster started in late 1980s, much later than its approval as a food additive (Buchman, Banerjee, & Hansen, 1988; Kaletta & Entian, 1989). The cluster contains genes for nisin biosynthesis and secretion: precursor gene NisA, dehydratase NisB, cyclase NisC, transporter NisT and protease NisP; it also contains immunity genes NisI and NisFEG, and regulatory genes NisRK (Kuipers, Beerthuyzen, Siezen, & de Vos, 1993; Qiao, et al., 1996) (Fig. 2.1 A). Nisin is synthesized ribosomally as a 57 amino acid precursor peptide, consisting of a 23-residue leader peptide and a 34-residue core peptide (Fig. 2.1 B). The presence of leader is crucial for the maturation enzymes to correctly process the core peptide, so that the 1 lanthionine (Lan), 4 methyllanthionines (MeLans), 1 dehydrobutyrine (Dhb) and 2 dehydroalanes (Dha) could be installed. NisB is responsible for dehydrating Ser/Thr to Dha/Dhb and NisC catalyzes the formation of thioether crosslinks between Cys and the Dha/Dhb, generating five intramolecular rings in the core region (Koponen, Tolonen, Qiao, Wahlstrom, Helin, & Saris, 2002) (Fig. 2.1 B). It was demonstrated that NisB and NisC form a multimeric membrane-associated complex with exporter NisT and that the dehydration and cyclization reactions intertwine (Siegers, Heinzmann, & Entian, 1996; Lubelski, Khusainov, & Kuipers, 2009), but the two enzymes are active independent of each other. The inactive mature nisin secreted by NisT is cleaved by the extracellular subtilisin-like Ser protease NisP to form active nisin (Qiao & Saris, 1996; van der Meer, Polman, Beerthuyzen, Siezen, Kuipers, & de

Vos, 1993) (Fig. 2.1 B). The producer cell is protected from its own product by the synergistic work of nisin-binding lipoprotein NisI and the ABC transporter NisFEG (Siegers & Entian, 1995; Qiao, Immonen, Koponen, & Saris, 1995; Stein, Heinzmann, Solovieva, & Entian, 2003; Takala, Koponen, Qiao, & Saris, 2004). Nisin acts as an auto-inducer and control its production through the constitutively expressed two-component regulatory system. The system is composed of a sensor histidine kinase NisK and a response regulator NisR. NisK senses the extracellular mature nisin and activates the transcription activator NisR. The activated NisR next activates the transcription of nisin biosynthesis genes and immunity genes (Engelke, Gutowski-Eckel, Kiesau, Siegers, Hammelmann, & Entian, 1994; Kuipers, Beerthuyzen, de Ruyter, Luesink, & de Vos, 1995; de Ruyter, Kuipers, Beerthuyzen, van Alen-Boerrigter, & de Vos, 1996).

The biosynthesis of nisin was largely studied *in vivo* with genetically modified producers. This approach was very successful in tracking the cellular location of the biosynthetic machinery, in probing the transient or permanent interactions between the enzymes and between the enzyme and the substrate, and in evaluating the tolerance of the whole system to the mutations in either substrate or enzymes. Nevertheless the tightly coupled system also hindered the study of each step in detail, such as the cofactor requirements of the dehydration and cyclization. Thus *in vitro* reconstitution of each step is the preferential approach to decouple the intertwined dehydration and cyclization reactions and study the enzymatic mechanisms. The cyclization reaction was first successfully reconstituted with heterologously produced dehydrated nisin precursor and the crystal structure of NisC was solved with bound zinc, which coordinates the thiolate of Cys in NisA (Li, Yu, Brunzelle, Moll, van der Donk, & Nair, 2006). The *in vitro* study of NisB, however, was only limited to its affinity to the non-processed or processed precursor peptides, although much is known about the peptide residues required for recognition

and NisB relaxed substrate specificity through *in vivo* studies (Kluskens, et al., 2005; Rink, de Boef, Leenhouts, Driessen, Moll, & Kuipers, 2005; Rink, et al., 2007; Plat, Kluskens, Kuipers, Rink, & Moll, 2011; Mavaro, et al., 2011; Khusainov, Moll, & Kuipers, 2013). The first successful *in vitro* reconstitution of NisB activity was accomplished recently by adding *E. coli* cell extract to the reaction. Also alanine scanning of conserved residues identified mutants that produce peptide intermediates modified with multiple Glu through ester bond (Garg, Salazar-Ocampo, & van der Donk, 2013). Fractionation of the *E. coli* cell extract confirmed RNA is key to the reaction as well as Glu. Further, Glu-tRNA^{Glu} is identified to participate the activation of Ser/Thr side chain through glutamylation and the subsequent elimination of Glu generates the double bond (Fig. 2.1 C). However, NisB shows no sequence similarity with other known proteins, which hinders further biochemical studies. To gain more insights on the recognition of both peptide and tRNA substrates and on the active sites of the glutamylation and elimination steps of dehydration reaction, I determined the crystal structure of NisB in the presence of its native substrate. Here, I present the 2.9 Å co-crystal structure of NisB in complex with peptide substrate NisA, which contains two separate domains for the two steps and NisA binds between the domains. The overall structure of NisB is unique but several smaller domains show similarity with structures deposited in the Protein Data Bank (PDB), which provide clues on their possible roles in the dehydration reaction. A docking model with bacterial tRNA^{Glu} is also established based on NisB structure.

Experimental Procedures

Cloning, Protein Expression and Purification

The *nisB* and *nisA* genes from nisin gene cluster of *Lactococcus lactis* sp. *lactis* were cloned and ligated into the MCS I and MCS II of pRSF-Duet vector (Novagen) respectively for co-expression. The resulting recombinant NisB contained an uncleavable N-termial hexahistitine tag (His-tag) while NisA was tagless. This plasmid pRSF-NisB-NisA was transformed into Rosetta 2 strain for expression. The culture first grew at 37 °C in LB medium supplemented with 50 μg/mL kanamycin and 25 μg/mL chloramphenicol until $O.D._{600}$ reached 0.5. Overexpression of NisB and NisA was induced with 0.3 mM IPTG at 18 °C for 20 h. Cells were pelleted by centrifugation at 3566 ×*g* for 25 min (SORVALL RC-3B Plus), resuspended with harvest buffer (20 mM Tris pH 8.0, 500 mM NaCl, 10% glycerol) and lysed with French Press (AVESTIN EmulsiFlex-C5). The supernatant was collected by centrifugation at 23645 ×*g* for 1 h (Beckman J2-21M/E) and loaded onto a 5mL HisTrap FF colume (GE Healthcare Life Sciences) pre-equilibrated with wash buffer (20 mM Tris, pH 8.0, 1 M NaCl, 5% glycerol, 30 mM imidazole). After washing with the same buffer for 5 column volumes to remove impurities, NisB-NisA complex was eluted with elution buffer (20 mM Tris, pH 8.0, 1 M NaCl, 5% glycerol, 200 mM imidazole) in a 0-100% linear gradient over 40 mL. Protein purity was checked with SDS-PAGE. Clean fractions were pooled and subjected to size exclusion chromatography with a HiLoad 16/600 Superdex 200 pg column (GE Healthcare Life Sciences) to remove aggregation and truncated protein. NisB-NisA complex was concentrated and underwent reductive methylation (Walter, et al., 2006). Briefly, 20 μL of 1 M dimethylaminoborane solution (DMAB) and 40 μL of 1 M formaldehyde solution were mixed with 1 mL protein solution followed by incubation at 4 °C for 2 h. The same solutions were added again and the incubation was repeated. After the final addition of 10 μL of 1 M DMAB, the mixture was incubated at 4 °C for 14 h before 100 μL of 1 M Tris pH 7.8 was added later to quench the reaction. The centrifuged

mixture was subjected to size exclusion chromatography with gel filtration buffer (20 mM HEPES pH 7.5, 300 mM KCl). The clean fractions were collected and concentrated for further use. The SeMet incorporated NisB-NisA protein was expressed and purified in a similar manner.

Crystallization and Structure Determination

Sitting drop vapor diffusion method was used for robotic crystallization condition screening and two initial hits were obtained (75 mM HEPES pH 7.5, 15% PEG10000, 25% (v/v) glycerol; 100 mM Bicine pH 8.5, 25% PEG6000). The initial hits were further optimized with hanging drop vapor diffusion method and only the second condition gave plate-like crystals suitable for diffraction data collection. A typical hanging drop was set up with 1 μL of NisB-NisA complex (5 mg/mL) mixed with 1 μL of precipitant solution (100 mM Bicine pH 8.5, 18-24% PEG6000) and was equilibrated over a well containing the same solution at 9 °C. Macro-seeding and micro-seeding facilitated the formation of crystals. SeMet NisB-NisA complex was crystallized in the same way. Diffraction-quality crystals were harvested around one week after seeding and stepwise equilibrated into crystallization condition supplemented with 10% (v/v) glycerol or 20% (v/v) ethylene glycol for protection before flash-frozen in liquid nitrogen.

Native and SeMet NisB-NisA data were collected at Sector 21 ID LS-CAT (Advanced Photon Source, Argonne National Labs, Illinois) and integrated and scaled using HKL2000 (Otwinowski & Minor, 1997) or XDS (Kabsch, 2010). Crystallographic phases were determined by single-wavelength anomalous diffraction method (SAD) with the 3.2 Å SeMet NisB-NisA data. Heavy atom sites were located using the SHELX suite (Thorn & Sheldrick, 2013), and refinement of heavy atom parameters in SHARP (Bricogne, Vonrhein, Flensburg, Schiltz, & Paciorek, 2003) yielded an initial figure of merit of 0.276. Density modification and two-fold

symmetry averaging produced experimental maps that permitted manual assignment of most secondary structural elements using COOT (Emsley & Cowtan, 2004). Further model building used the 2.9 Å resolution data collected on crystals of native NisB-NisA. Although obvious electron density corresponding to the NisA peptide can be observed in experimentally phased maps, the peptide was only built into the model once the free R factor for the structure had decreased to below 0.30. The validity of all models was routinely determined with MOLPROBITY (Chen, et al., 2010) and by using the free R factor to monitor improvements during building and crystallographic refinement.

Results

Overall structure of NisB-NisA complex

In order to investigate the structural basis of the dehydration reaction and substrate recognition, we determined the 2.9 Å crystal structure of NisB in complex with its peptide substrate NisA. The 117-kDa NisB forms a dimer in the asymmetric unit (Fig. 2.2 A), consistent with the previous report in which the experimentally predicted NisB molecular weight in aqueous solution corresponds to a dimeric composition. The two protomers in the asymmetric unit clutch each other and face the opposite directions. In each protomer, residues S5–I706 constitute the N-terminal large domain and residues R716–E990 compose the C-terminal small domain (Fig. 2.2 A). The two domains form a cleft and NisA peptide binds in the cleft (Fig. 2.2 B). Residues spanning I706–R716 and V893–G919 were not modeled in the structure due to the lack of electron density.

Mutations in NisB that have been demonstrated to interfere with either glutamate addition (glutamylation reaction) or elimination reactions cluster separately into the N-terminal domain or

C-terminal domain in the structure (Fig. 2.3 A). Mutations that abolished the glutamylation reaction but retained *in vitro* glutamate elimination activity on glutamylated NisA intermediates (Garg, Salazar-Ocampo, & van der Donk, 2013) locate within a 10 Å-radius region in the N-terminal domain (Fig. 2.3 B). This domain is therefore designated glutamylation domain. Likewise, mutations of R786, R826 and H961 that led to accumulation of glutamylated peptide intermediates and thus were important for elimination activity (Garg, Salazar-Ocampo, & van der Donk, 2013) cluster in a small region in the C-terminal domain (Fig. 2.3 C), together with other residues crucial for NisB dehydration process (Garg, Salazar-Ocampo, & van der Donk, 2013). This domain is consequently designated elimination domain. The regions harboring the catalytically crucial residues are proposed to be the glutamylation site and elimination site, respectively. Interestingly, the partitioned two domains correspond to two discrete proteins of around 800 residues and 300 residues encoded in thiopeptide biosynthetic gene clusters (Li & Kelly, 2010), which indicates that the dehydration of thiopeptide is achieved in a similar manner – aminoacylation and elimination catalyzed by these two proteins.

tRNA Binding Cavity of NisB

RNA-binding proteins often contain positively charged patches on the accessible surface to mediate electrostatic interactions with their polyanionic cognate RNAs (Shazman & Mandel-Gutfreund, 2008). The surface electrostatic potential calculation on NisB reveals a densely positively charged cavity on the inner surface of the glutamylation domain (Fig. 2.4 A). The dimensions of this cavity could suitably accommodate the characteristic L shape of a typical tRNA. Based on the limited structural similarity of this region to the *Xenopus laevis* double-stranded RNA-binding protein A complexed with its cognate RNA (PDB 1DI2), a docking

model was proposed for NisB binding to a bacterial $tRNA^{Glu}$, using the structure of $tRNA^{Glu}$ from *Thermus thermophilus* (PDB 1N78) (Fig. 2.4 B). The binding of tRNA to the positively charged cavity in NisB is mainly through the acceptor stem, which is for amino acid attachment. This suggests glutamylation domain possibly recognizes the acceptor stem instead of the overall L shape of tRNA. Similar interactions have been observed in peptidyltransferases that utilize aminoacyl tRNA for nonribosomal peptide synthesis, such as the peptidyltransferase FemX involved in the synthesis of the pentapeptide in peptidoglycan precursor (Mainardi, Villet, Bugg, Mayer, & Arthur, 2008). FemX contains a positively charged grove that reaches to the catalytic site (Biarrotte-Sorin, Maillard, Delettré, Sougakoff, Arthur, & Mayer, 2004) and its activity depends on correct recognition of the acceptor stem of the tRNA donor, demonstrated by both biochemical and structural studies (Fonvielle, et al., 2009; Fonvielle, et al., 2013). Although NisB catalyzes entirely different reaction from FemX and has a much more complicated structure, it could use a similar strategy to bind the tRNA substrate. Moreover, the docking model also shows the tRNA acceptor stem is placed in proximity to the glutamylation site, as a further support for the model.

Structural Homology of NisB

Structural similarity search of NisB against Protein Data Bank (PDB) gave no obviously homologous structure, but searching with smaller domains gave several structurally similar domains or proteins. The domain spanning residues F136–I216 in glutamylation domain resembles the N-terminal domain of LynD (and TruD) (Fig. 2.5 A), the heterocyclase involved in cyanobactins biosynthesis (Koehnke J. , et al., 2013; Koehnke, et al., 2015). It contains a three-stranded antiparallel β sheet followed by a three-α helical bundle and is named winged helix-

turn-helix motif (wHTH). wHTH typically recognizes DNA with its alpha helix in the binding event of transcription factors (Brennan, 1993). In the case of NisB or LynD, this domain recognizes the conserved N-terminal leader peptide or recognition sequence (RS) of the peptide substrate. Residues from both α helical bundle and a β strand provide interactions with the peptide. Leader peptide or RS binds the helical bundle of wHTH with its N-terminus and stretches out as a β strand to join the antiparallel β sheet. Although catalyzing distinct chemical reactions in the biosynthesis of different RiPP classes, NisB and LynD do share a common feature as RiPP processing enzymes – the scheme for peptide binding, suggesting a more universal role of this domain in enzyme-peptide substrate interaction.

A α/β subdomain spanning F734–Y820 in the elimination domain of NisB is structurally homologous to the protein LsrG in bacterial interspecies quorum sensing system (Fig. 2.5 A). LsrG catalyzes epimerization of processed autoinducer-2 molecule by abstracting a proton adjacent to a carbonyl group (Marques, et al., 2011). Despite their similar structure that consists a four-stranded β sheet flanked by α helices, the active site of LsrG is not conserved in NisB elimination domain. The proposed active site residues of LsrG are placed between the β sheet and α helices based on the observation of electron density of an unidentified ligand in the structure (Fig. 2.5 B). However, the residues crucial for NisB glutamate elimination are located between the LsrG-like subdomain and the second subdomain of the elimination domain (Fig. 2.5 C). Scrutinization of this region led to the identification of an additional conserved basic residue in the pocket, R784. Alanine substitution of R784 resulted in the accumulation of glutamylated peptide intermediates as predicted, further supporting the assignment of the active site for the elimination reaction.

Peptide Binding Site of NisB

Although NisB is co-purified with the full-length NisA, only continuous electron density of residues K-20 to K-9 of NisA leader peptide was clearly observed in the peptide-binding groove between the glutamylation domain and the elimination domain (Fig. 2.6). The electron density diminishes towards the C-terminus of the leader peptide and the entire core peptide, possibly due to their flexible conformation and transient interaction with NisB. The NisA leader peptide forms a β strand and interacts antiparallely with the β strand of the peptide-binding domain, spanning residues L166–N175 (Fig. 2.2 A). The location of the leader-binding groove and the extended conformation of the leader peptide ensure that the core peptide, instead of any residue in the leader region, can flexibly reach both the distant glutamylation site and elimination site. Conserved in other class I lantibiotics, NisA contains a F-D/N-L-N/D motif essential for NisB recognition (Mavaro, et al., 2011; Plat, Kluskens, Kuipers, Rink, & Moll, 2011; Khusainov, Moll, & Kuipers, 2013). The F-18 of the motif is situated in a hydrophobic cage formed by NisB residues V176, V198, Y202, L209 and Y213 (Fig. 2.6). The L-16 interacts with NisB hydrophobically in a pocket composed of I171, Y213 and L217 (Fig. 2.6). The D-15 of the conserved motif forms hydrogen bond with R154 of NisB (Fig. 2.6). The extensive interactions between NisA FNLD motif and NisB in the crystal structure support and explain the observation that mutations in this motif greatly decrease the binding affinity of NisA to NisB.

Discussion

As one of the longest known lantibiotic, nisin has been under intensive investigation for decades, with respect to its *in vivo* production, regulation, mode of action, and industrial application. However, the detailed biosynthetic process wasn't elucidated until recently,

especially the synthesis of dehydrated amino acids. NisB was assigned as the dehydratase based on *in vivo* experiments, but it was the successful *in vitro* reconstitution of NisB activity and the decoupling of glutamylation and glutamate elimination reactions that significantly advance the understanding and promote the further illustration of the tRNA dependent mechanism. Here, we reported the 2.9 Å crystal structure of NisB-NisA complex, the first structure of class I lantibiotic dehydratase. The glutamylation and elimination reactions are isolated from each other spatially in two separate active sites in glutamylation domain and elimination domain, respectively. The peptide substrate is positioned in the peptide-binding groove between the two domains, allowing for reaching either of the active site and proper movement between the two sites. The flexibility of NisA core peptide could explain the diminished electron density for the NisA C-terminus in the structure. In contrast, the leader peptide is tightly held by the leader peptide-binding domain of NisB, mainly through hydrophobic interactions with the NisA FNLD motif. The retainment of this highly conserved motif in class I lantibiotic precursor peptides suggests their cognate dehydratases use a similar mechanism for peptide recognition. That the NisB leader peptide-binding domain is isolated from catalytic sites provides explanations for the observation that Ser/Thr residues in the leader peptide are never dehydrated and that a minimal distance is required between the FNLD motif and the first dehydrated amino acid (Plat, Kluskens, Kuipers, Rink, & Moll, 2011). The leader peptide is protected by the leader-binding domain in the structure and the minimal distance required is the distance between this motif and the glutamylation site. This distance also justifies the general N-to-C directionality and distributive manner of NisA dehydration reactions catalyzed by NisB: residues closer to the leader peptide are more likely to access the glutamylation site than the C-terminal residues.

As the representative LanB protein, the structure of NisB in complex with NisA shed lights on the catalytic mechanism and substrate recognition of other LanBs in the biosynthesis of respective class I lantibiotics. Based on sequence similarity of class I LanBs, the cognate Glu-tRNA^{Glu} dependent two-step dehydration mechanism is likely a prevalent strategy. The discrete glutamylation site, elimination site and the leader peptide-binding motif are possibly also conserved. Further study on the mechanism and structure of other LanBs will be necessary to confirm this prediction. Besides, LanB gene homologs are also widely identified in the biosynthetic gene clusters of other RiPP classes, such as in the aforementioned thiopeptide gene clusters. The glutamylation and elimination domains of NisB are encoded as two separate polypeptides but the two-step reaction strategy probably also applies here to achieve dehydrated amino acids. Whether these enzymes prefer Glu-tRNA^{Glu} or other amino acids, however, is worth investigating.

Aminoacyl-tRNA has been shown to participate diverse chemical transformations and cellular processes apart from the canonical ribosomal polypeptide synthesis (Francklyn & Minajigi, 2010). In the biosynthesis of amino-acid-containing natural products, it often serves as the donor of activated amino acid and peptide bond or isopeptide bond is formed to attach the amino acid as part of the final product. In contrast, the NisB dehydration reaction utilizes the activated amino acid as the replacement of ATP to activate another amino acid and bring in a "good" leaving group. An ester bond rather than peptide bond is formed to ensure the subsequent elimination together with proton abstraction. This novel usage of aminoacyl-tRNA for amino acid activation is an addition to its diverse functions other than ribosomal polypeptide synthesis. Nevertheless, although a tRNA-binding site is observed on NisB and a binding model is proposed, structures of the NisB-NisA-GlutRNA^{Glu} or NisB-NisA-acceptor stem trimeric

complex will be desired to thoroughly evaluate the NisB conformational changes in different stages of the catalytic cycle, and to illustrate the interactions between the acceptor stem and the tRNA-binding site, between the aminoacylated 3’ end and the glutamylation active site, and between the glutamyl group and the Ser/Thr of NisA core peptide.

Figures

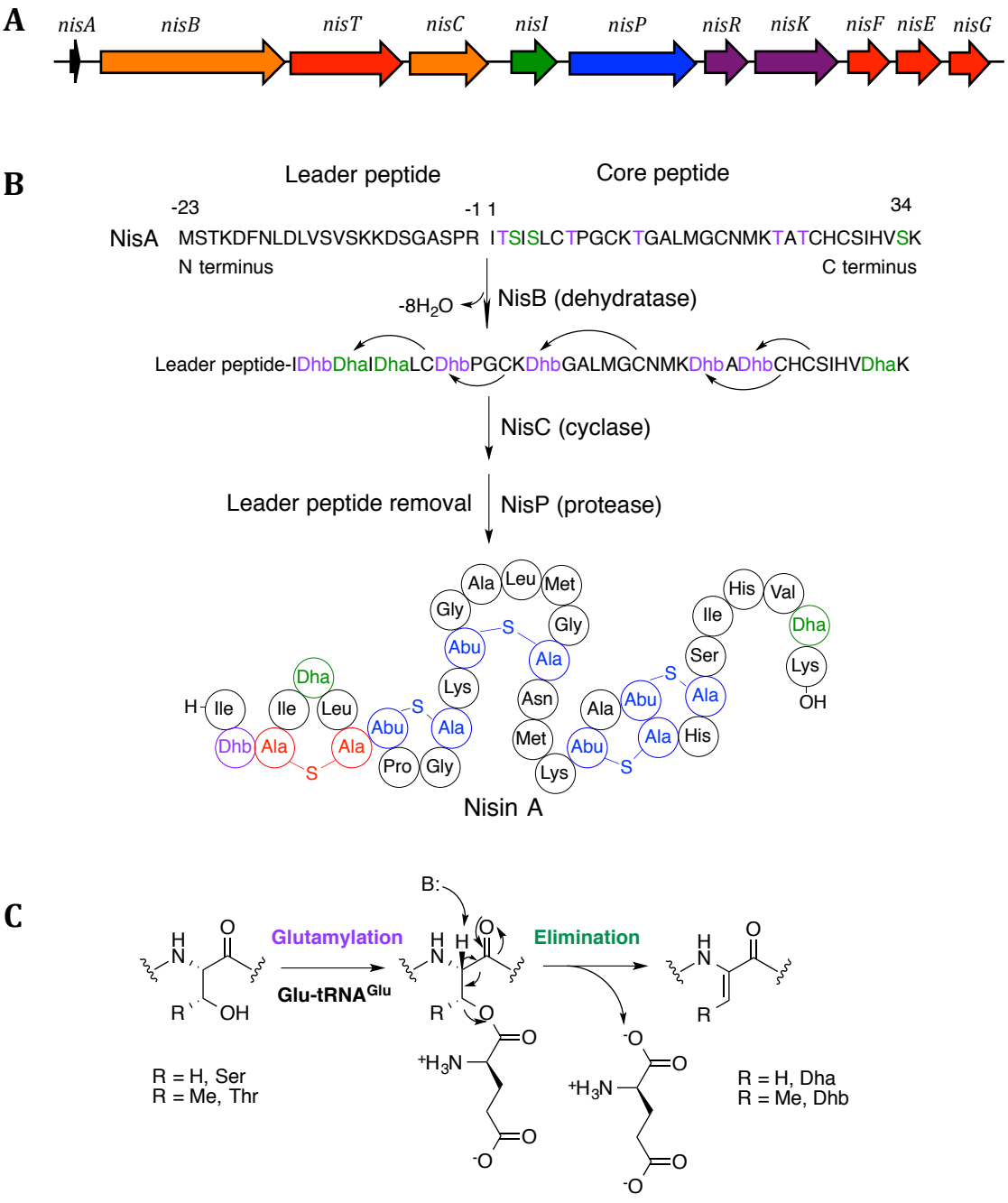

Fig. 2.1 **A.** Nisin biosynthetic gene cluster from *Lactococcus lactis*. *nisA* encodes the precurspr peptide; *nisB* and *nisC* encode the dehydratase and cyclase, respectively; *nisT* encodes the

transporter for nisin secretion and *nisP* encodes the extracellular protease; *nisRK* encode the two-component regulatory system; *nisI* and *nisFEG* are immunity genes. **B.** The biosynthetic pathway of nisin. Ribosomally synthesized NisA precursor peptide contains leader peptide and core peptide. The latter undergoes post-translational modifications catalyzed by dehydratase NisB and cyclase NisC. Ser (green) is converted to Dha and Thr (purple) is converted to Dhb by NisB. NisC catalyzes the nucleophilic attack of Cys side chain onto the Dha and Dhb to form Lan (red) and MeLan (blue). The subsequent leader peptide removal generates the mature nisin. **C.** Mechanism of dehydration reaction. NisB uses Glu carried on Glu-tRNA^{Glu} to activate Ser and Thr. The side chain hydroxyl group attacks the ester bond of Glu-tRNA^{Glu} and gets glutamylated. The subsequent elimination of Glu by NisB produces dehydrated amino acid.

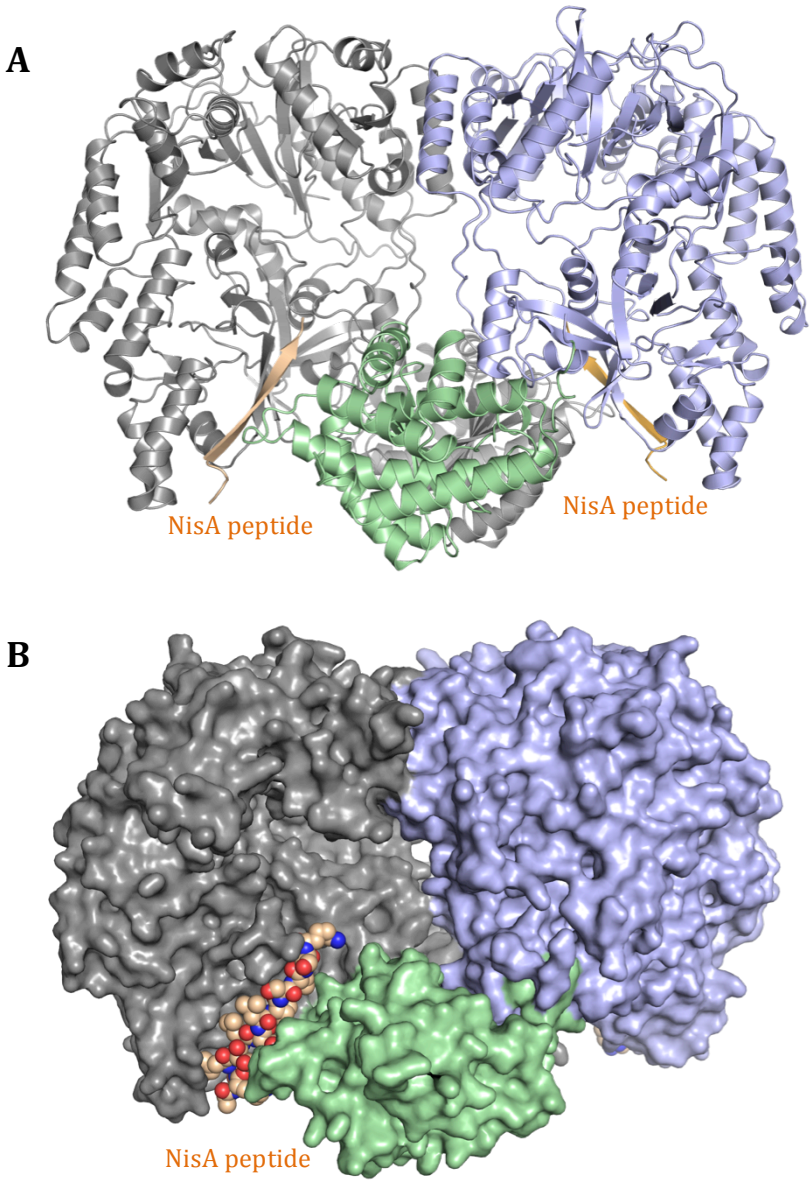

Fig. 2.2 **A.** Overall structure of NisB in complex with peptide substrate NisA. NisB forms a dimer in the crystal structure. One protomer is in gray; the other protomer is colored differently with two domains. The N-terminal large domain is in purple and the C-terminal small domain is in green. NisA peptide is colored in gold. **B.** Surface representation of NisB dimer. NisA peptide is shown as spheres with oxygen atoms colored in red and nitrogen atoms colored in blue.

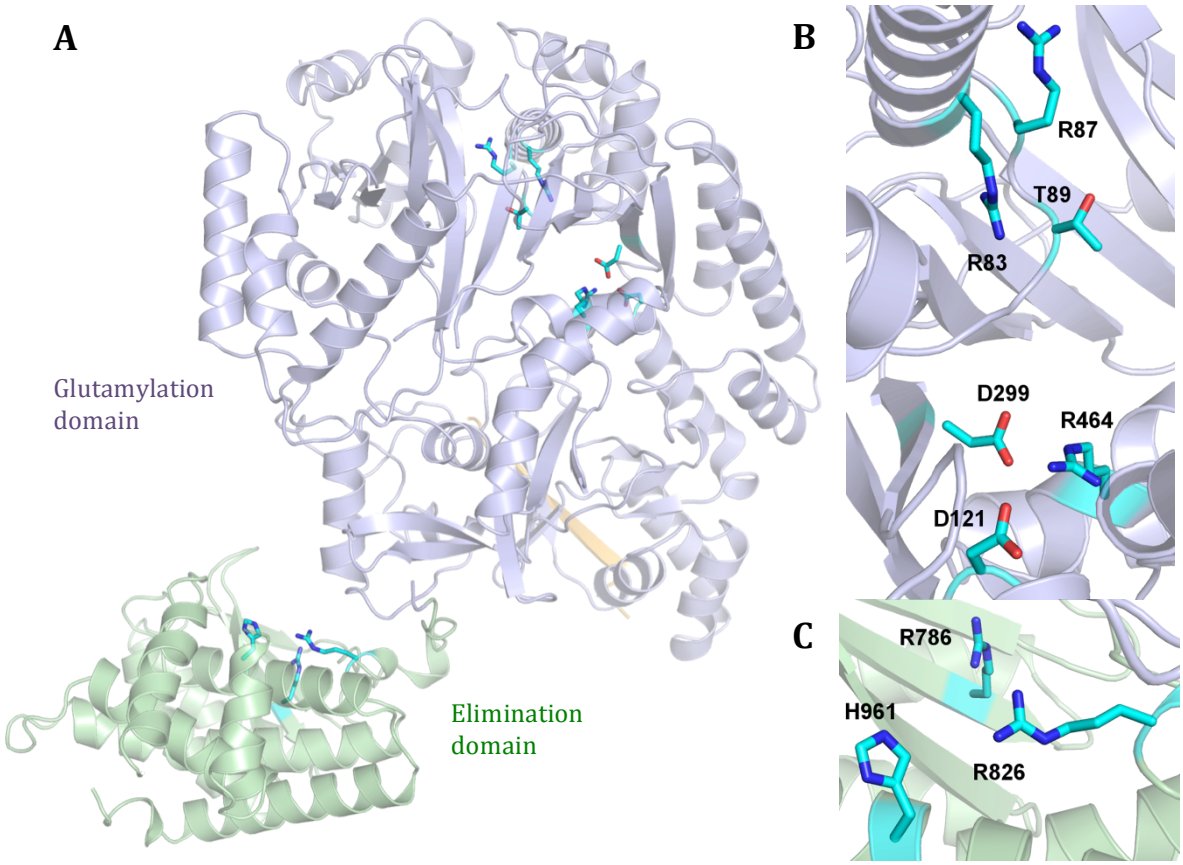

Fig. 2.3 **A.** A protomer of NisB with corresponding NisA substrate. The N-terminal glutamylation domain is colored in purple and the C-terminal elimination domain is colored in green. NisA peptide is shown as a β strand in gold. The residues essential for either glutamylation reaction or elimination reaction are clustered separately and shown as sticks. **B.** A close view of residues important for Ser/Thr glutamylation reaction. **C.** A close view of residues important for glutamate elimination reaction.

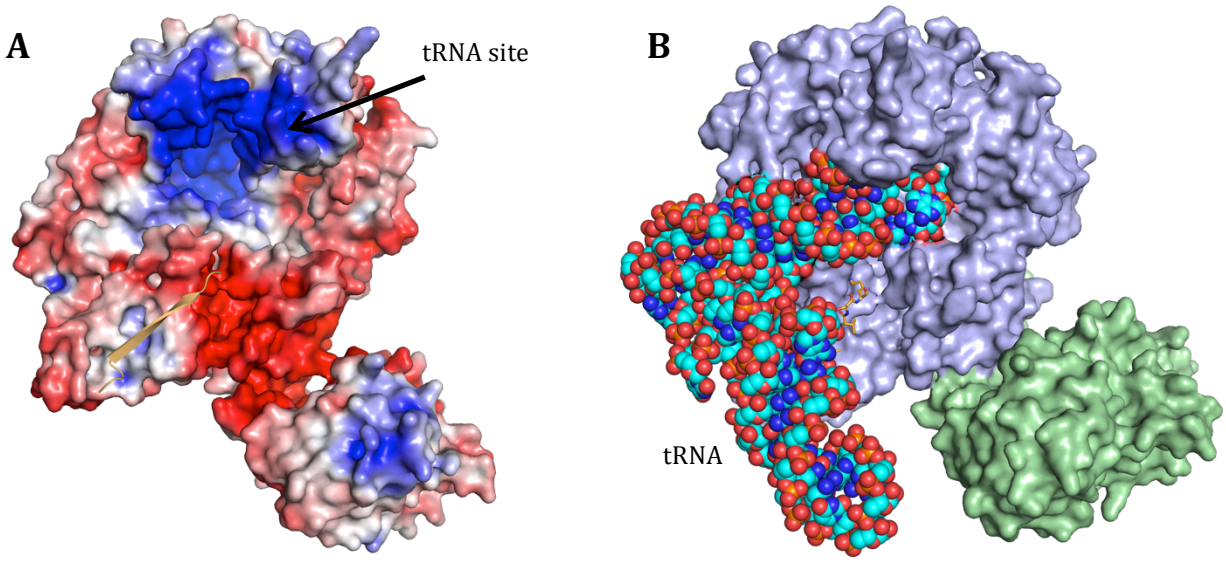

Fig. 2.4 **A.** Surface representation of a NisB protomer showing surface electrostatics. Blue colored region represents positively charged surface; red colored region represents negatively charged surface. The saturation of the color is positively correlated with the density of the charge. Electroneutral surface is colored white. The highly basic region in glutamylation domain is proposed as the tRNA-binding site. NisA peptide is shown as a β strand in gold. **B.** A docking model of bacterial $tRNA^{Glu}$ onto NisB. The NisB glutamylation domain is colored purple and the elimination domain is in green. The tRNA is shown as spheres and NisA peptide is shown as sticks. The acceptor stem of tRNA is bound at the proposed tRNA-binding site and points to the region with residues important for glutamylation reaction.

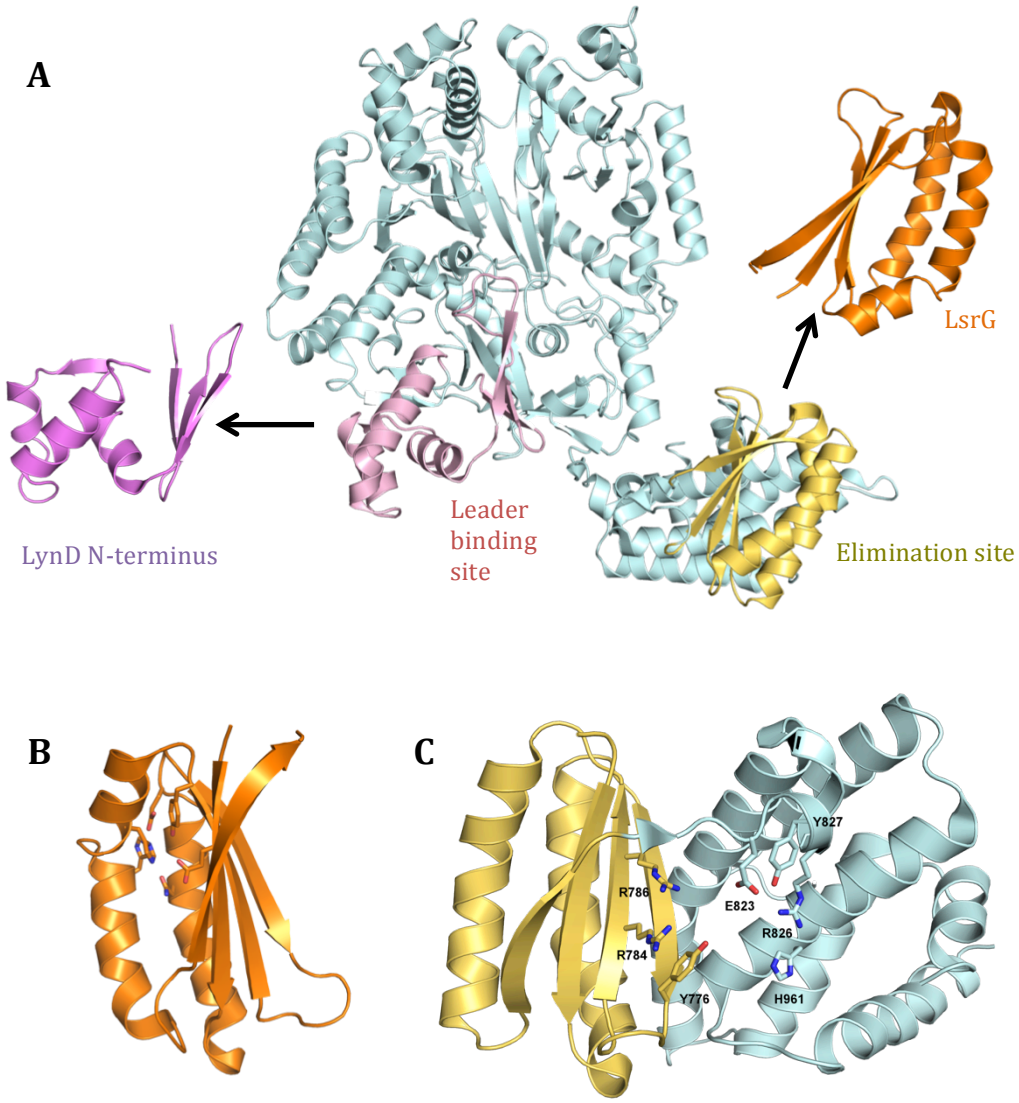

Fig. 2.5 **A.** Structure homology of NisB. Homology search with a NisB protomer or either of the two domains showed no similar structure in PDB. Smaller domains in NisB are structurally similar to known proteins. The leader-binding domain of NisB is homologous to the LynD N-terminal domain, which is also responsible for binding the recognition sequence of the peptide substrate. A small domain in the elimination domain is homologous to LsrG involved in degradation of quorum-sensing molecule. **B.** Structure of LsrG with the proposed active-site residues shown in sticks. Based on the electron density of an unidentified substrate in the LsrG

crystal structure, the active site is proposed to be between the β sheet and α helices. **C.** The elimination domain of NisB. It is divided into two domains and the α/β domain homologous to LsrG is colored in yellow. Compared to the proposed LsrG active site, the active site for NisB elimination reaction is located between the two domains and composed of residues from both domains (shown as sticks).

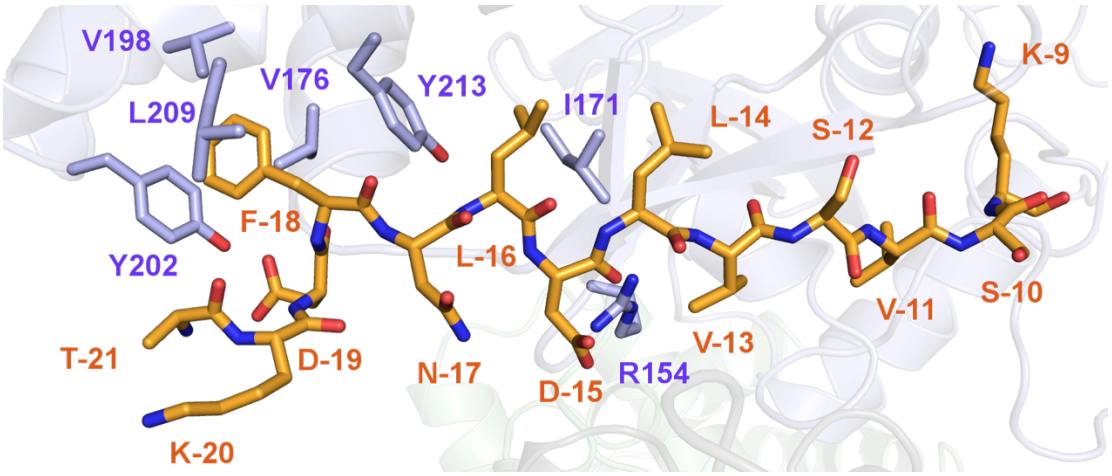

Fig. 2.6 NisA leader peptide coordination in NisB leader-binding site. The residues K-20 through K-9 (shown as sticks in orange) observed in the co-crystal structure are coordinated through hydrophobic interactions and hydrogen bonds. NisB residues I171, V176, V198, Y202, L209 and Y213 form a hydrophobic cage to accommodate the F-18 and L-16 in the conserved FNLD motif in the leader peptide. NisB residue R154 interacts with D-15 through hydrogen bond. The leader peptide forms a β strand antiparallel to the β strand in the leader-binding site and hydrogen bonds through backbone atoms also assist in substrate coordination.

2.2 Structural Study of Lantibiotic Dehydratase MibB in Microbisporicin Biosynthesis

Introduction

Infections caused by Gram-positive pathogens represent a global challenge and multidrug-resistant strains keep emerging. An effective strategy in treating Gram-positive bacteria is to disrupt cell wall biosynthesis. Lantibiotics are promising chemotherapeutic agents in inhibiting peptidoglycan synthesis, especially for the resistant strains, and they show no cross-resistance with glycopeptides (Cotter, Hill, & Ross, 2005). Microbisporicin is a class I lantibiotic discovered recently through high-throughput screening of uncommon actinomycete broth extracts for peptidoglycan biosynthesis inhibitors except β-lactams and glycopeptides. With the commercial name NAI-107, it is in late preclinical-phase trials and demonstrated potent inhibitory activity against a panel of Gram-positive pathogenic bacteria, including methicillin-resistant *Staphylococcus aureus* (MRSA) and *vancomycin-resistant Enterococci* (VRE), with comparable or superior efficacy with respect to other antibiotics tested such as lantibiotic nisin and vancomycin (Castiglione, et al., 2008; Jabés, Brunati, Candiani, Riva, Romanó, & Donadio, 2011). And it is more active in treating Gram-negative bacteria *Moraxella catarrhalis*, *Neisseria* spp. and *Haemophilus influenzae*. Similar to the well studied class I lantibiotic nisin, it binds cell wall precursors such as lipid I and lipid II and preferentially forms a 2:1 complex. It disrupts cell membrane and depolarizes it, but unable to form pores on the membrane to cause acute cessation of DNA, RNA and protein biosynthesis (Münch, et al., 2014).

Microbisporicin contains 3 lanthionine (Lan), 1 methyllanthionine (MeLan) crosslinks and a C-terminal S-((*Z*)-2-aminovinyl)-D-cysteine (AviCys), which prevents C-terminal hydrolysis. It also has two unusual modifications: 5-chlorotryptophan and 3,4-dihydroxylproline, which are

unique among lantibiotics (Fig. 2.7 B). There is not much known about microbisporicin biosynthesis, but the general biosynthetic route can be predicted from its gene cluster (Fig. 2.7 B). The cluster was identified and annotated in *Microbispora corallina* lately and was shown to successfully produce active microbisporicin heterologously (Fig. 2.7 A). It encodes precursor peptide MibA, maturation enzymes for installing all functional groups and for secretion, immunity proteins and transcription regulators for feed-forward regulation (Foulston & Bibb, 2010; Foulston & Bibb, 2011). MibA is a 57-residue precursor peptide with the first 33 amino acids as leader peptide. The leader contains the conserved "FNLD" recognition motif (LDLD in MibA) and a typical leader cleavage site. Same as other class I lantibiotics, MibB and MibC are encoded separately and show sequence similarities with representative LanB and LanC proteins, respectively (Foulston & Bibb, 2010).

The LanB protein in nisin biosynthesis, NisB, has recently been characterized as a tRNA-dependent dehydratase and catalyzes both glutamylation and elimination steps. Apart from its cognate Glu-tRNA^{Glu}, NisB is capable of using *E. coli* Glu-tRNA^{Glu} for the reaction with both *in vitro* and *in vivo* evidence (Garg, Salazar-Ocampo, & van der Donk, 2013; Ortega, Hao, Zhang, Walker, & van der Donk, 2015). It is intriguing to evaluate the prevalence of this biosynthetic scheme in class I lantibiotic biosynthesis, as well as in other biosynthetic pathways involving LanB homologs. MibB, encoded by a rare actinomycete, serves as a suitable candidate to address questions such as whether utilization of tRNA is a common strategy for LanBs, whether a specific variant of Glu-tRNA is preferred and whether *E. coli* Glu-tRNA could be recognized and function universally. As part of the assessment, I determined the MibB structure at 2.7 Å without peptide substrate. MibB appears as a dimer in the crystal structure and is composed of two domains similar to those in the previously determined NisB structure. The residues crucial

for glutamylation reaction and elimination reaction in NisB are also conserved in MibB structure and mapped separately in the two domains. Close to MibB putative glutamylation site, there is a highly basic region potentially for tRNA recognition. MibB also contains a α+β domain for leader peptide binding as in NisB. However due to the poor solubility of peptide substrate, attempt on co-crystalling MibB-MibA complex was not successful.

Experimental Procedures

Cloning, Protein Expression and Purification

The *mibB* gene from microbisporicin gene cluster of *Microbisppora corallina* was cloned into pET-28b vector (Novagen) between *Nde*I and *Xho*I sites. The resulting plasmid was transformed into Rosetta 2 strain for expression. The culture first grew at 37 °C in LB medium supplemented with 50 μg/mL kanamycin and 25 μg/mL chloramphenicol until $O.D._{600}$ reached 0.5. Overexpression of MibB was induced with 0.3 mM IPTG at 18 °C for 20 h. Cells were pelleted by centrifugation at 3566 ×*g* for 25 min (SORVALL RC-3B Plus), resuspended with harvest buffer (20 mM Tris pH 8.0, 500 mM NaCl, 10% glycerol) and lysed with French Press (AVESTIN EmulsiFlex-C5). The supernatant was collected by centrifugation at 23645 ×*g* for 1 h (Beckman J2-21M/E) and loaded onto a 5 mL HisTrap FF colume (GE Healthcare Life Sciences) pre-equilibrated with wash buffer (20 mM Tris, pH 8.0, 1 M NaCl, 5% glycerol, 30 mM imidazole). After washing with the same buffer for 5 column volumes to remove impurities, MibB was eluted with elution buffer (20 mM Tris, pH 8.0, 1 M NaCl, 5% glycerol, 200 mM imidazole) in a 0-100% linear gradient over 40 mL. Protein purity was checked with SDS-PAGE. Clean fractions were pooled and subjected to size exclusion chromatography with a HiLoad 16/600 Superdex 200 pg column (GE Healthcare Life Sciences) to remove aggregation

and truncated protein. MibB was concentrated and underwent reductive methylation (Walter, et al., 2006). Briefly, 20 μL of 1 M dimethylaminoborane solution (DMAB) and 40 μL of 1 M formaldehyde solution were mixed with 1 mL protein solution followed by incubation at 4 °C for 2 h. The same solutions were added again and the incubation was repeated. After the final addition of 10 μL of 1 M DMAB, the mixture was incubated at 4 °C for 14 h before 100 μL of 1 M Tris pH 7.8 was added to quench the reaction. The centrifuged mixture was subjected to size exclusion chromatography with gel filtration buffer (20 mM HEPES pH 7.5, 300 mM KCl). Clean fractions were collected and concentrated for further use.

Crystallization and Structure Determination

A sitting drop vapor diffusion method was used for robotic crystallization condition screening. The initial hit (100 mM HEPES pH 7.5, 20% (v/v) ethanol) was further optimized with hanging drop vapor diffusion method. A typical hanging drop was set up with 2 μL of MibB (7 mg/mL) mixed with 2 μL of precipitant solution (100 mM HEPES pH 7.5 or 100 mM MOPS pH 7.9, 16-22% (v/v) ethanol) and was equilibrated over a well containing the same solution at 16 °C. Rod-like diffraction-quality crystals were harvested around one week after setting up and stepwise equilibrated into crystallization condition supplemented with 20% (v/v) ethylene glycol for protection before flash-frozen in liquid nitrogen.

Crystallographic phases were determined by multiple isomorphous replacement method (MIR) using heavy atom derivative with potassium hexachloroosmate (IV). Native and derivative MibB data were collected at Sector 21 ID LS-CAT (Advanced Photon Source, Argonne National Labs, Illinois) and integrated and scaled using HKL2000 (Otwinowski & Minor, 1997) or XDS (Kabsch, 2010). Heavy atom sites were located using the SHELX suite

(Thorn & Sheldrick, 2013), and refinement of heavy atom parameters in SHARP (Bricogne, Vonrhein, Flensburg, Schiltz, & Paciorek, 2003) yielded an initial figure of merit of 0.274. Density modification and two-fold symmetry averaging produced experimental maps that permitted manual assignment of most secondary structural elements using COOT (Emsley & Cowtan, 2004). A higher-resolution data set was collected at Station F at CHESS (Cornell University, New York) and integrated and scaled using XDS. Refinement of the initial model against this data set allowed for a near complete model to be manually built. The validity of all models was routinely determined with MOLPROBITY (Chen, et al., 2010) and by using the free R factor to monitor improvements during building and crystallographic refinement.

Results

In order to gain structure insights of the putative dehydratease MibB and the dehydration reaction it catalyzes, we determined the 2.7 Å crystal structure of MibB. MibB crystallize as a dimer in the asymmetric unit with two protomers clutching each other and facing the opposite directions (Fig. 2.8). The 120-kDa protomer is divided into a large N-terminal domain and a small C-terminal domain, forming a cleft between the two domains. Residues D4–S809 compose the N-terminal domain and residues P819–H1114 comprise the C-terminal domain. Residues E20 through D78 were not modeled into the structure due to the lack of electron density.

The overall structure of MibB is homologous to the recently characterized structure of class I lantibiotic dehydratase NisB although they share only 20% sequence identity. The two dimeric structures superimpose decently with a RMSD of 3.5 Å (Fig. 2.9). Since NisB exists as a dimer in the aqueous solution and crystalizes as a dimer, the structurally homologous MibB may also function as a dimer, which suggests the possibility that the homodimer formation is essential for

the reaction catalyzed by class I LanB enzymes. MibB also shares notable structural features of NisB that are essential and characteristic for a tRNA-dependent lantibiotic dehydratase and indicate a similar two-step dehydration mechanism.

Primary sequence alignment of MibB and NisB reveals that the residues essential for NisB glutamylation and elimination reactions are all conserved in MibB. Similar to NisB, the two groups of residues cluster in the N-terminal domain and C-terminal domain respectively (Fig. 2.10). Corresponding MibB residues that are important for glutamylation reaction in NisB locate within a 10 Å-radius region in MibB N-terminal domain. This domain is therefore designated glutamylation domain (Fig. 2.11 A, B). Likewise, corresponding MibB residues that are important for glutamate elimination reaction in NisB cluster in a small region in the C-terminal domain. This domain is thus designated glutamate elimination domain (Fig. 2.11 A, C). The regions harboring the conserved catalytically crucial residues are proposed to be the glutamylation site and elimination site as in the case of NisB.

Similar to NisB, the region close to the glutamylation site is structurally homologous to the *Xenopus laevis* double-stranded RNA-binding protein A (PDB 1DI2) and is highly positively charged on the surface (Fig. 2.11 D). The dimensions of this region are close to the ones of the proposed tRNA-binding pocket on NisB, and thus it might as well be responsible for aminoacyl-tRNA substrate recognition by interacting with the acceptor stem. A small domain spanning residues Q201–L286 in MibB glutamylation domain resembles the leader peptide-binding domain identified in NisB (Fig. 2.11 E, F). Although the MibB structure presented here is not in complex with its peptide substrate MibA, this domain is predicted to contribute to MibA leader peptide binding from structure homology. With the presence of the conserved LDLD motif in MibA leader peptide, similar hydrophobic and hydrogen-bonding patterns between this motif and

the MibB leader peptide-binding domain are expected based on the observed interactions in NisB-NisA complex structure (Fig. 2.12 A, B). Compared to the NisB leader-binding domain bound with NisA leader, the MibB leader-binding domain demonstrates an obvious conformational change (Fig. 2.12 C). The α helical bundle is shifted outwards with respect to the β sheet, suggesting less engagement of the residues comprising the hydrophobic cage in a leader-free state. Notably, the utilization of homologous leader peptide-binding domain for substrate recognition is observed beyond the class I lanthipeptides. The heterocyclases LynD and TruD in the biosynthesis of cyanobactins, another class of RiPPs, contain a similar fold at the N-terminus and have been shown structurally to recognize the corresponding recognition sequence in peptide substrate. Despite the drastically different protein structures and reactions they catalyze, the lantibiotic dehydratases and cyanobactin heterocyclases do depend on leader peptide (or recognition sequence) recognition for catalysis, suggesting a general role of this scaffold in proper peptide binding in RiPP biosynthesis.

Discussion

Recent biochemical and structural studies on lantibiotic dehydratase NisB revealed that the dehydrated amino acids of class I lantibiotics are installed by NisB in a Glu-tRNA^{Glu} dependent manner and through glutamylation and glutamate elimination steps. To evaluate the universality of this dehydration mechanism in class I lantibiotic biosynthesis, we provided the structural support by determining the 2.7 Å crystal structure of MibB, the putative dehydratase involved in the biosynthesis of the newly discovered class I lantibiotic microbisporicin produced by an uncommon actinomycete. MibB shows high structural homology with NisB, including the key features that support a tRNA-dependent two-step dehydration mechanism. As with NisB, MibB

consists of the N-terminal glutamylation domain and C-terminal glutamate elimination domain with the conserved catalytically important residues grouping in a small region of the respective domains. Primary sequence alignment of MibB, NisB, with SpaB involved in subtilin biosynthesis reveals that these important residues are highly conserved in lantibiotic dehydratases (Fig. 2.10), indicating that the two-step strategy of introducing dehydrated amino acids is probably universal in class I lantibiotic biosynthesis. A similar RNA-binding fold with positively charged surface is identified near the MibB glutamylation site and proposed to recognize the Glu-tRNA^{Glu} based on the comparison with NisB. However, while NisB doesn't distinguish between the *E. coli* Glu-tRNA^{Glu} and its cognate Glu-tRNA^{Glu} as glutamate donor, MibB strictly depends on its cognate Glu-tRNA^{Glu} for activity. Despite the expected electrostatic interactions between NisB/MibB and the tRNA acceptor stem, additional recognition mechanisms are believed to participate in determining the tRNA selectivity, such as the interactions between specific residues in the tRNA-binding pocket and the bases of the acceptor stem at specific positions. To answer this question, a structure of NisB/MibB in complex with the cognate Glu-tRNA^{Glu} or with acceptor stem analogs will be desired to thoroughly elucidate the recognition mechanism.

Despite the lack of substrate MibA peptide in the MibB structure, a small domain homologous to the leader peptide-binding domain of NisB is observed at the boundary of the MibB glutamylation domain and elimination domain. Predicted from the leader peptide recognition pattern in NisB, this domain is the leader peptide-binding domain of MibB and recognizes MibA leader peptide in a similar manner. In addition, the location of this domain would preferentially allow the movement of MibA core peptide between the glutamylation and elimination active sites. As mentioned above, the same structural fold also serves as the

recognition sequence-binding domain of the heterocyclase involved in cyanobactin biosynthesis. A recent study demonstrated that this domain is actually widely distributed in the leader peptide-dependent biosynthetic enzymes of diverse RiPP classes (Burkhart, Hudson, Dunbar, & Mitchell, 2015), indicating a general strategy for peptide substrate recognition. However, it remains elusive but is intriguing to investigate how prevalent this strategy is among different RiPP modifying enzymes and whether additional structural scaffolds are also used for leader peptide recognition.

An apparent difference between the MibB leader-binding domain and that of NisB is the movement of the α helical bundle with respect to the β sheet. This conformational change could be explained by several possibilities. It might represent the intrinsic arrangement of structural elements in MibB that is distinct from NisB, even though they utilize similar strategy for leader peptide recognition. A second possible explanation for the conformational change is that it might result from the absence of substrate peptide, which signals the peptide-binding domain to stay in a less engaged state. The binding of substrate peptide could possibly introduce a conformational change and the α helical bundle therefore moves closer to the β sheet to clutch the peptide substrate. To interpret this observation properly, structure of MibB in complex with its cognate substrate MibA is desired for comparison.

An interesting phenomenon is the similar dimeric structure of NisB and MibB, two class I LanB enzymes from distinct bacterial phyla. First, this similarity indicates that the dimeric structure is likely to be universal for class I lantibiotic dehydratases despite the drastic differences of the source organism. Second, the dimer formation is possibly a prerequisite for proper catalysis by LanB enzymes. It is obvious in both MibB and NisB structures that the distance for a peptide substrate to reach the elimination site of the same protomer is much longer

than that to the corresponding elimination site of the other protomer. It could be explained by a possible *trans* manner of the glutamate elimination step. The NisA peptide is glutamylated in the protomer it is located in but might take a shorter path and gets eliminated in the elimination site of the other protomer. To further investigate this possibility, a co-crystal structure of LanB with its cognate LanA locked in the elimination site would be necessary.

Figures

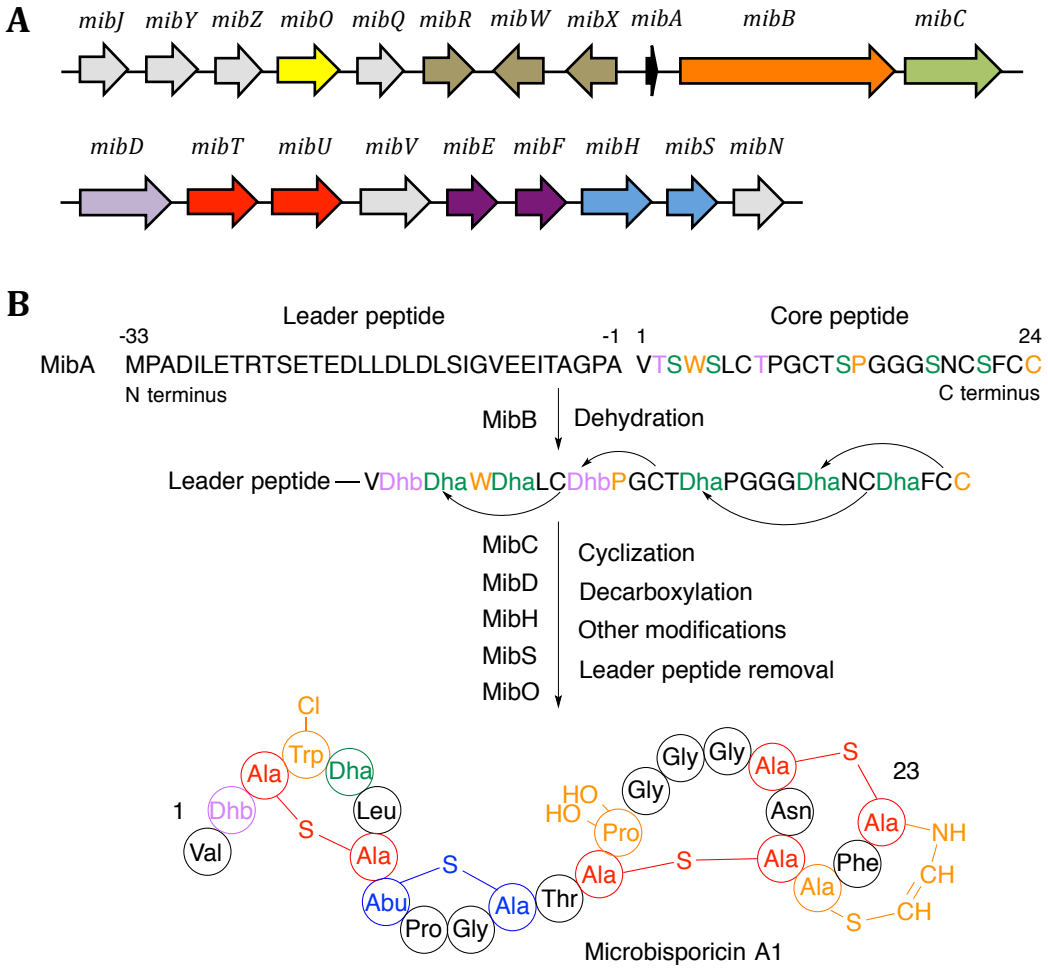

Fig. 2.7 **A.** Microbisporicin biosynthetic gene cluster from *Microbispora corallina*. *mibA* encodes precursor peptide; *mibB* and *mibC* encode lantibiotic dehydratase and cyclase, respectively; *mibD* encodes a putative flavoprotein to decarboxylate the C-terminus and catalyze AviCys formation; *mibHS* encode a putative flavin-dependent halogenase for Trp chlorination; *mibO* encodes a putative P450 for Pro hydroxylation; *mibTU*, ABC transporters; *mibEF*, immunity genes; the product of *mibX*, *mibW*, *mibR* are involved in transcriptional regulation. **B.** Biosynthetic pathway of microbisporicin A1.

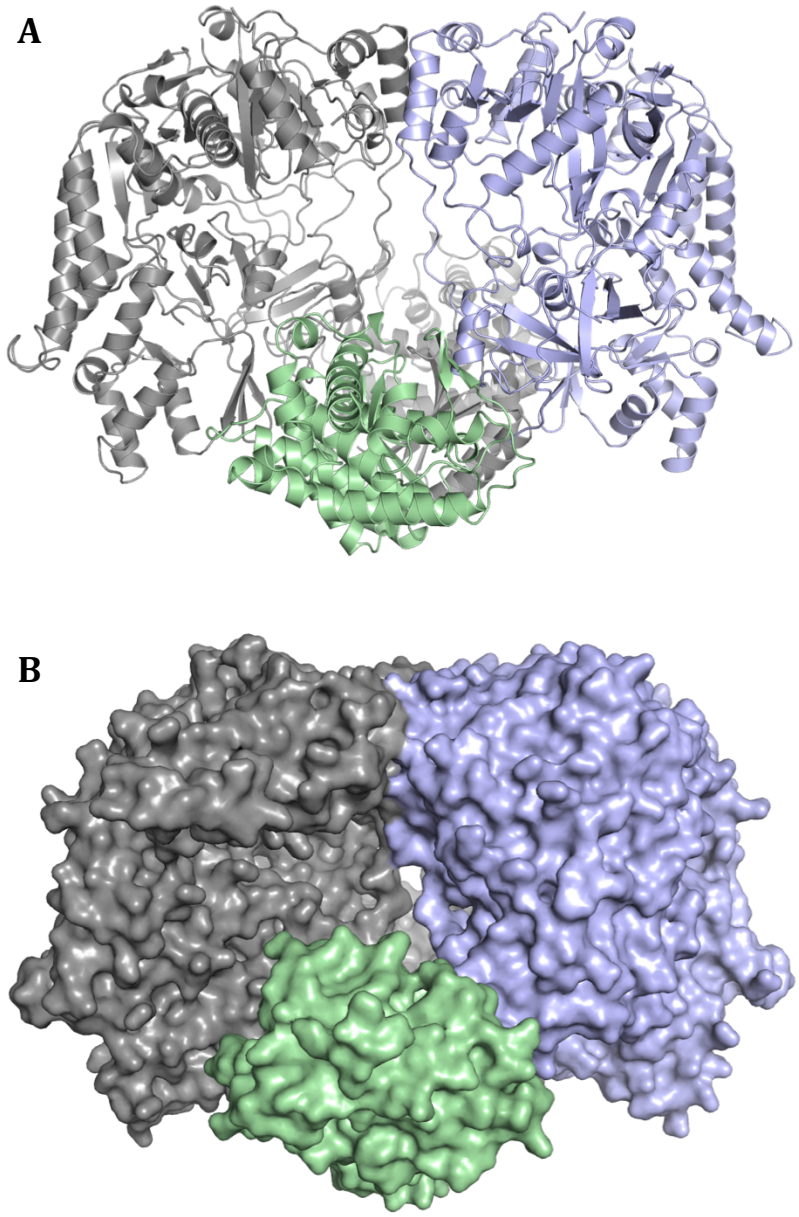

Fig. 2.8 **A.** Overall structure of MibB. MibB forms a dimer in the crystal structure. One protomer is in gray; the other protomer is colored differently with two domains. The N-terminal large domain is in purple and the C-terminal small domain is in green. **B.** Surface representation of MibB dimer.

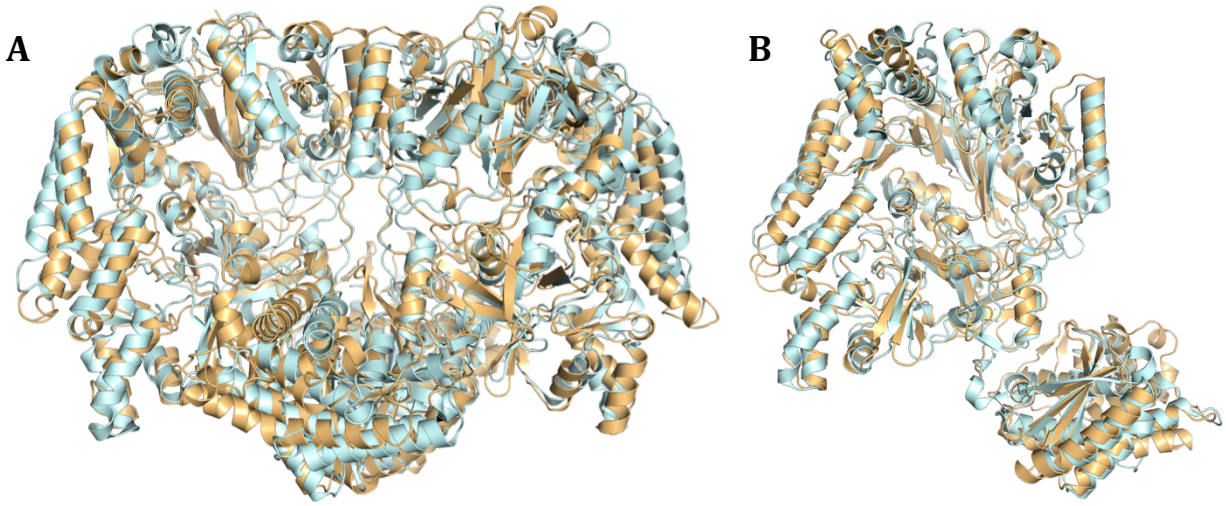

Fig. 2.9 **A.** Superimposition of MibB dimer with NisB dimer. MibB is colored in gold and NisB is colored in cyan. **B.** Superimposition of MibB protomer with NisB protomer. There are no substantial structural differences between the two class I lantibiotic dehydratases from different phyla of bacteria.

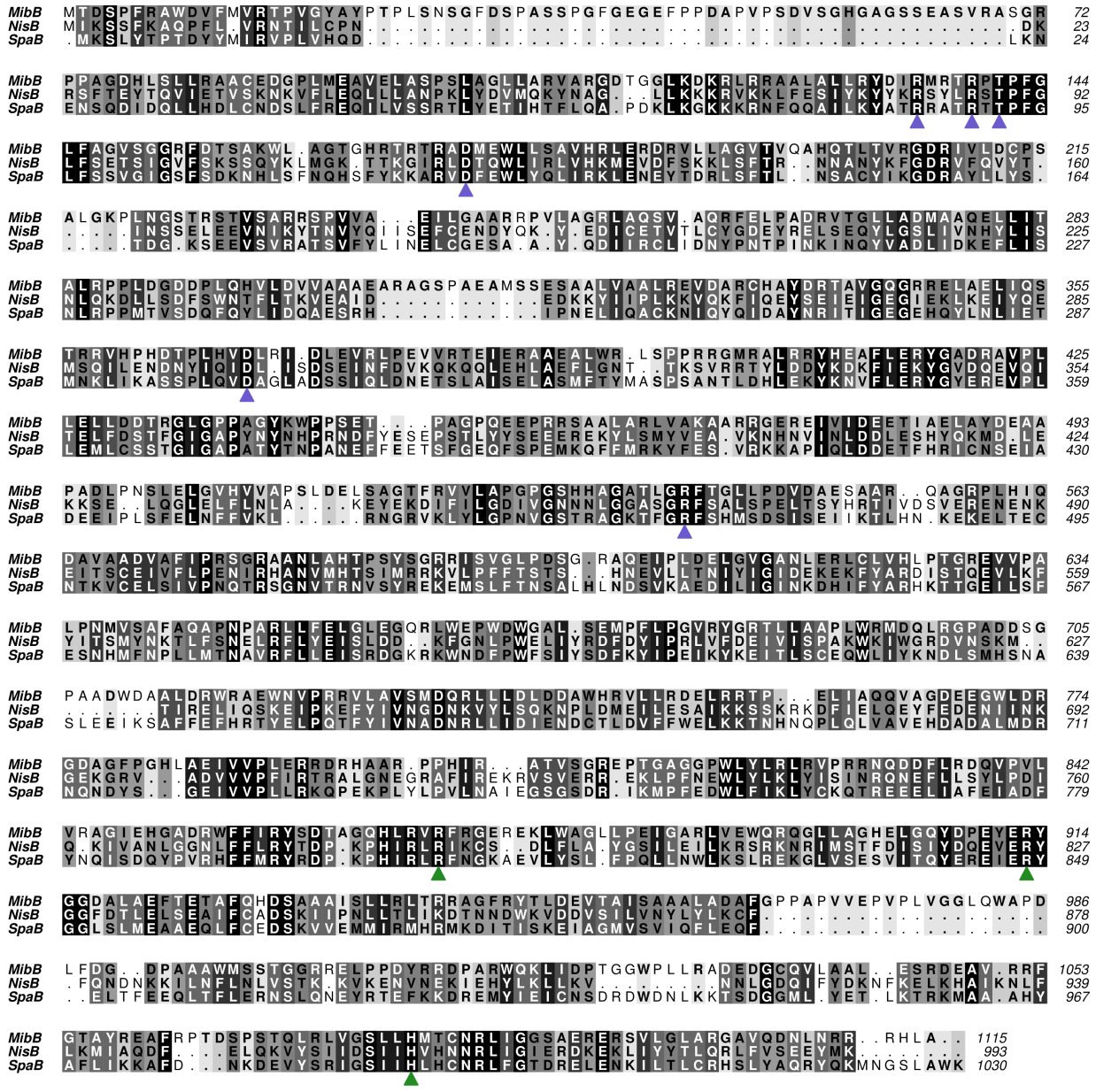

Fig. 2.10 Sequence alignment of three class I lantibiotic dehydratases MibB, NisB and SpaB. Residues important for glutamylation reaction and elimination reaction are highly conserved and indicated with purple and green arrowheads, respectively.

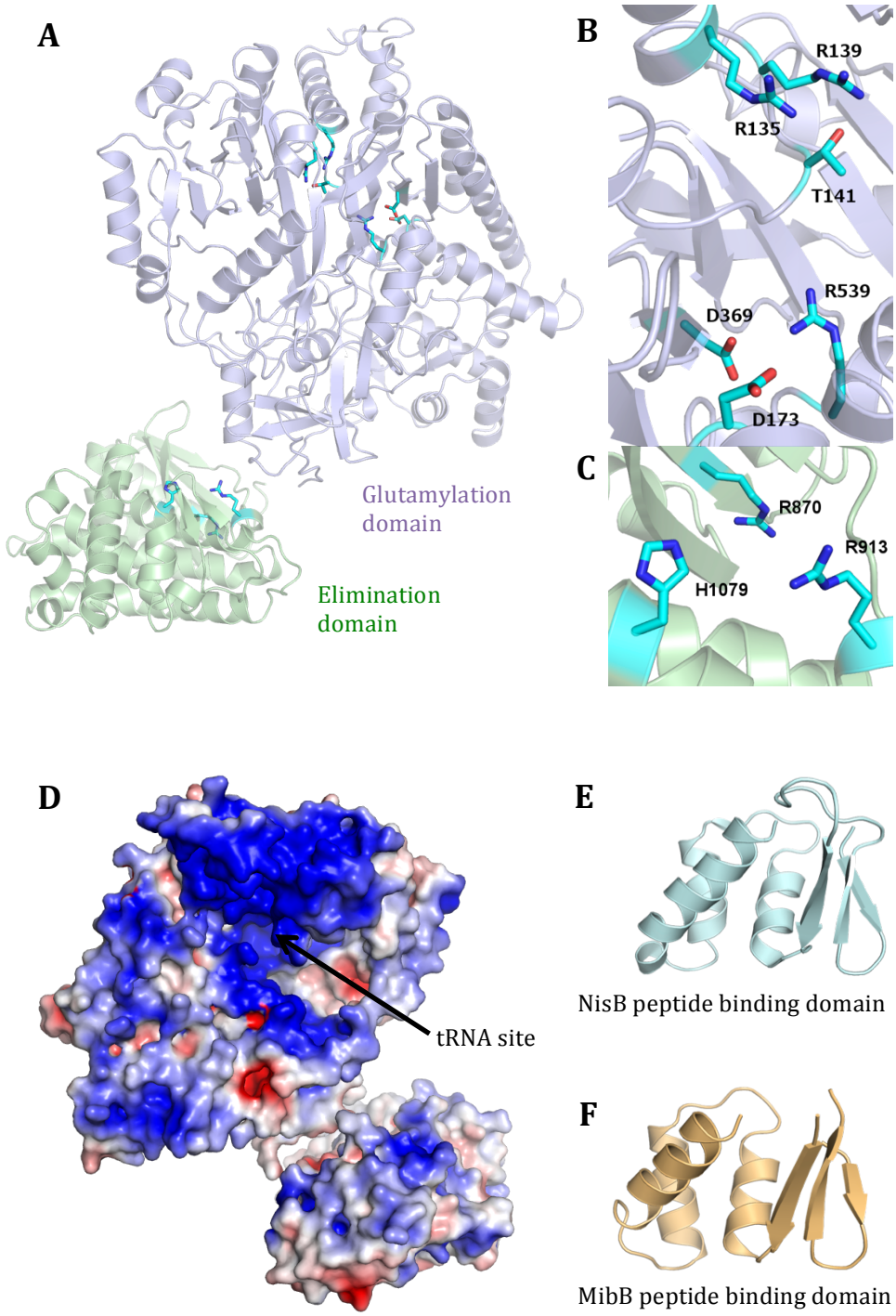

Fig. 2.11 **A.** A protomer of MibB with glutamylation domain (purple) and elimination domain (green). The residues essential for either glutamylation reaction or elimination reaction are clustered separately into the two domains and shown as sticks (cyan). **B.** A close view of

conserved residues important for Ser/Thr glutamylation reaction. **C.** A close view of conserved residues important for glutamate elimination reaction. **D.** Surface representation of MibB protomer showing surface electrostatics. Blue colored region represents positively charged surface; red colored region represents negatively charged region. The saturation of the color is positively correlated with the density of the charge. Electroneutral surface is colored white. The highly basic region in glutamylation domain is proposed as the tRNA-binding site. **E.** The NisB leader-binding domain spanning residues A147–H221. **F.** The MibB domain spanning residues L202–E279 is proposed as the leader peptide-binding domain, based on structural homology with the NisB leader-binding domain.

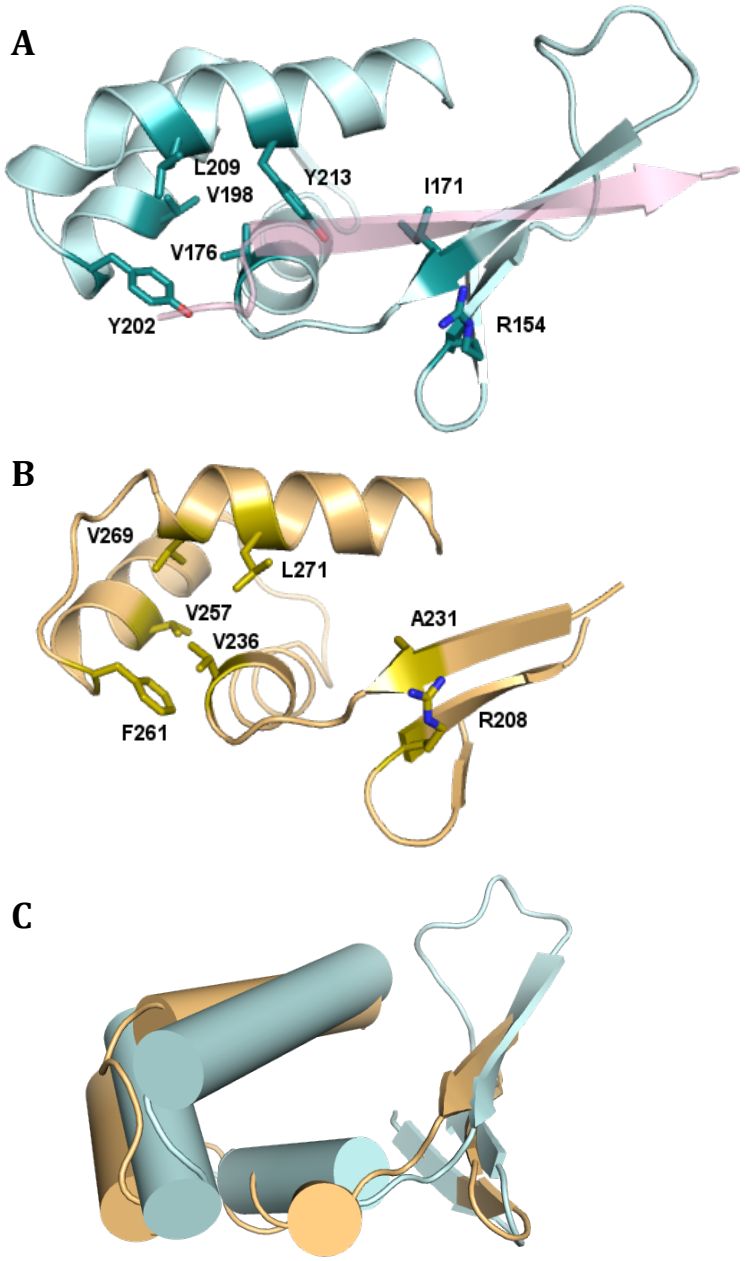

Fig. 2.12 **A.** NisB leader-binding domain (cyan) bound with NisA leader peptide (pink). Residues involved in hydrophobic interactions and hydrogen-bonding interactions with the conserved NisA FNLD motif are shown as sticks. **B.** Proposed MibB leader-binding domain (gold). MibA contains a LDLD motif in the leader peptide. Despite the lack of MibA in the crystal structure, residues potentially important for LDLD motif recognition are identified based

on the structural homology with NisB leader-binding domain. **C.** Superimposition of MibB leader-binding domain (gold) and NisB leader-binding domain (cyan). An outward shift of the α helices is observed with respect to the β sheet in MibB leader-binding domain compared with the NisB leader-binding domain.

Chapter 3 Structural and Biochemical Studies on the *N*-Methyltransferases in Plantazolicin Biosynthesis[1]

Introduction

Plantazolicin (PZN) was first isolated from the Gram-positive plant growth-promoting rhizobacterium *Bacillus amyloliquefaciens* FZB42 and selectively inhibits the growth of closely related *Bacilli*, including the anthrax-causing bacterium *Bacillus anthracis* (Scholz, et al., 2011; Molohon, et al., 2011). PZN is a linear azol(in)e-containting peptide (LAP). The 41-residue precursor peptide is synthesized on the ribosome and contains a 27-residue long leader peptide and a Cys/Ser/Thr rich core peptide (Fig. 3.1 B). The Cys/Ser/Thr residues are clustered into two adjacent patches and converted to two poly-azol(in)e moieties by the BCD tailoring enzyme complex (Kalyon, et al., 2011; Molohon, et al., 2011). The heterocyclase D catalyzes ATP-dependent cyclodehydration reaction and azoline rings form between the side chain of Cys/Ser/Thr and the preceding amide bond. The dehydrogenase B then oxidizes the azolines to corresponding aromatic azoles. A total of 10 azolines are installed with 5 in each patch and the first 9 of them are oxidized, leaving the last methyloxazoline ring susceptible to hydrolysis. The leader peptide is cleaved by a zinc dependent protease, yielding the rigid desmethylPZN. The last post-translational modification that confers the bioactivity is the leader-independent dimethylation of the N-terminal Arg (Fig. 3.1 B). The desmethylPZN with unmodified N-

[1] Part of this chapter is adapted from the following published articles:

Lee, J., Hao, Y., Blair, P. M., Melby, J. O., Agarwal, V., Burkhart, B. J., Nair, S. K., & Mitchell, D. A. (2013). Structural and functional insight into an unexpectedly selective N-methyltransferase involved in plantazolicin biosynthesis. *Proceedings of the National Acadmy of Sciences of the United States of America, 110* (32), 12954-12959.

Hao, Y., Blair, P. M., Sharma, A., Mitchell, D. A., & Nair, S. K. (2015). Insights into methyltransferase specificity and bioactivity of derivatives of the antibiotic plantazolicin. *ACS Chemical Biology, 10* (5), 1209-1216.

terminal α-amine completely loses its activity against *Bacillus anthracis* (Molohon, et al., 2011). A SAM-dependent methyltransferase L (BamL) identified in the PZN gene cluster of *Bacillus amyloliquefaciens* FZB42 is proposed to catalyze the dimethylation reaction on the peptide substrate (Fig. 3.1 A).

The dimethylation of the N-terminal α-amine of a ribosomal peptide is unusual in bacteria. The only other examples observed to date are in the biosynthesis of linaridins, including cypemycin and grisemycin (Claesen & Bibb, 2010; Claesen & Bibb, 2011). Both of them contain N^{α}, N^{α}-dimethyl-alanine at N-terminus and the lack of this modification also abolishes the bioactivity (Claesen & Bibb, 2011). Biochemical studies on the *N*-methyltransferase CypM in cypemycin biosynthesis showed that CypM is very promiscuous towards peptide substrate. It dimethylates the α-amine of small polar or nonpolar amino acids at N-terminus, and also the ε-amine of an exposed Lys in the peptide (Zhang & van der Donk, 2012). In contrast to CypM, BamL requires the presence of both the N-terminal Arg and the following polyazolic structure to be catalytically competent. The substitution of Arg with homoarginine, ornithine, Lys or other representative amino acids didn't trigger methylation reaction (Piwowarska, Banala, Overkleeft, & Süssmuth, 2013). The same goes for the replacement of the polyazole with small nonpolar residues, noncyclized precursors, or the heterocyle mimic Pro (Lee, et al., 2013). The single successful attempt in generating substrate analog was to keep the Arg and shorten the length of the polyazolic structure. BamL could effectively dimethylate the synthetic desmethylPZN analogs with 5 azoles (pentazole), 3 azole (triazole) and even 1 azole (monoazole) (Piwowarska, Banala, Overkleeft, & Süssmuth, 2013). It also works on Arg amide, but only under single turnover condition.

The stringent substrate requirement makes us wonder what are the structural differences between BamL and typical methyltransferases and what are the key elements in determining substrate recognition. To answer these questions, a structure of BamL in complex with its cognate substrate is desired. However, the attempt on co-crystallization was not successful due to the poor aqueous solubility of desmethylPZN. Obtaining substrate affinity to BamL or kinetic parameters for the methylation reaction was also impeded. Fortunately the more soluble monoazolic, triazolic and pentazolic desmethylPZN analogs (Arg-Az_1, Arg-Az_3 and Arg-Az_5, respectively) (Fig. 3.1 C) were synthesized and showed low micromolar binding affinity to the enzyme (Sharma, Blair, & Mitchell, 2013). These analogs were used for co-crystallization with BamL as well as BpumL, the PZN methyltransferase from *Bacillus pumilus* ATCC7061, which shares 48% sequence identity with BamL. Here I reported the high-resolution structures of BamL in complex with *S*-adenosylhomocysteine (SAH) alone or in complex with both SAH and Arg-Az_1/Arg-Az_3. I also presented the co-crystal structures of BpumL with SAH or with SAH and Arg-Az_1/Arg-Az_3/Arg-Az_5. These truncated analogs also facilitated the kinetic parameter determination of both wide-type enzyme and active site mutants designed based on structure information.

Experimental Procedures

Cloning, Protein Expression and Purification

The *bamL* gene from plantazolicin gene cluster of *Bacillus amyloliquefacians* FZB42 was cloned and ligated into pET-MBP vector, which is modified from commercial pET-28b vector (Novagen). The resulting recombinant BamL was fused to the C-terminus of the maltose binding protein (MBP) and a tobacco etch virus protease (TEV) recognition site between them is used for

cleavage. This plasmid pET-MBP-BamL was transformed into Rosetta 2 strain for expression. The culture first grew at 37 °C in LB medium supplemented with 50 μg/mL kanamycin and 25 μg/mL chloramphenicol until $O.D._{600}$ reached 0.5. Overexpression of MBP-BamL was induced with 0.3 mM IPTG at 18 °C for 20 h. Cells were pelleted by centrifugation at 3566 ×*g* for 25 min (SORVALL RC-3B Plus), resuspended with harvest buffer (20 mM Tris pH 8.0, 500 mM NaCl, 10% glycerol) and lysed with French Press (AVESTIN EmulsiFlex-C5). The supernatant was collected by centrifugation at 23645 ×*g* for 1 h (Beckman J2-21M/E) and slowly loaded onto a 20 mL amylose column (NEB) pre-equilibrated with harvest buffer. After washing with the same buffer for 5 column volumes to remove impurities, MBP-BamL was eluted directly with MBP elution buffer (harvest buffer supplemented with 20 mM maltose). Protein purity was checked with SDS-PAGE and clean fractions were pooled. Protein solution was then treated with TEV protease at 4 °C overnight with the addition of 0.5 mM dithiothreitol (DTT). Protein solution was further loaded onto a 5 mL HisTrap FF colume (GE Healthcare Life Sciences) pre-equilibrated with harvest buffer. After washing with the same buffer for 2 column volumes, the tag-free BamL was eluted with wash buffer (20 mM Tris, pH 8.0, 1 M NaCl, 5% glycerol, 30 mM imidazole). The His-tagged MBP could be removed from the column with elution buffer (20 mM Tris, pH 8.0, 1 M NaCl, 5% glycerol, 200 mM imidazole). The BamL fractions were concentrated and the buffer was changed back to harvest buffer while concentrating.

The *bpumL* gene from plantazolicin gene cluster of *Bacillus pumilus* ATCC7061 was cloned and ligated into pET-28b vector between *Nde*I and *Xho*I sites. The resulting recombinant BpumL contained a cleavable N-terminal hexahistidine tag (His-tag). The plasmid transformation, heterologous expression and cell disruption procedures were the same as pET-MBP-BamL. The supernatant was loaded onto a 5 mL HisTrap FF column. After extensive washing with wash

buffer, the His-BpumL was eluted with elution buffer in a 0-100% linear gradient over 40 mL. The protein purity was checked by standard SDS-PAGE and the purer fractions were collected and subjected to tag cleavage by thrombin. The tag-free BpumL was further purified by size exclusion chromatography with a HiLoad 16/600 Superdex 75 pg column (GE Healthcare Life Sciences). The protein was stable in gel filtration buffer (20 mM HEPES pH 7.5, 100 mM KCl) and was concentrated for further use. SeMet incorporated BpumL was expressed and purified in a similar manner.

Crystallization and Structure Determination

A sitting drop vapor diffusion method was used for robotic crystallization condition screening. The initial hits were further optimized with hanging drop vapor diffusion method. A typical hanging drop was set up with 1 μL of protein-substrate complex mixed with 1 μL of precipitant solution and was equilibrated over a well containing the same solution at 4 °C (for BamL) or 9 °C (for BpumL). The concentrations used for BamL and BpumL were 10 mg/mL and 8 mg/mL, respectively. The SAM and SAH were purchased from Sigma and used at 2 mM for crystallization. The plantazolicin analogs Arg-Az_1, Arg-Az_3, and Arg-Az_5 were courtesy of Prof. D. Mitchell lab (Sharma, Blair, & Mitchell, 2013) and were incubated with enzymes at 2 mM prior to crystallization. Crystals of BamL-SAH grew in 100 mM NaOAc pH 4.6, 1.8-2.2 M $(NH_4)_2SO_4$. Crystals of BamL-Arg-Az_1-SAH grew in 100 mM MES pH 6.5, 200 mM $(NH_4)_2SO_4$, 26-32% PEG5000MME. Crystals of BamL-Arg-Az_3-SAH grew in 100 mM Bis-tris pH 5.5, 200 mM Li_2SO_4, 18-24% PEG3350. Crystals of BpumL or SeMet-incorporated BpumL with SAH grew in 100 mM HEPES pH 7.5, 200 mM NaCl, 18-24% PEG3350. Crystals of BpumL-Arg-Az_1-SAH grew in 100 mM Tris pH 8.5, 200 mM $MgCl_2$, 24% PEG4000. Crystals

of BpumL-Arg-Az_3-SAM grew in 100 mM Tris pH 8.5, 200mM $MgCl_2$/NaOAc, 20%-24% PEG3350/PEG4000/PEG8000. Crystals of BpumL-Arg-Az_5-SAM grew in 100 mM Tris pH 8.5, 200 mM $MgCl_2$, 24-27% PEG4000. All the crystals were stepwise equilibrated into crystallization condition supplemented with 10% (v/v) glycerol for protection before flash-frozen in liquid nitrogen.

All of the data sets were collected at Sector 21 ID LS-CAT (Advanced Photon Source, Argonne National Labs, Illinois) and integrated and scaled using HKL2000 (Otwinowski & Minor, 1997) or XDS (Kabsch, 2010). Initial crystallographic phases were determined by single-wavelength anomalous diffraction method (SAD) with a five-fold redundant data set of SeMet BpumL. Heavy atom sites were located using the HySS, and refinement of heavy atom parameters in SHARP (Bricogne, Vonrhein, Flensburg, Schiltz, & Paciorek, 2003) yielded an initial figure of merit (FOM) of 0.293 to 3.2 Å resolution. Solvent flattening using DM further improved the quality of the initial map (FOM = 0.665). Although the map was of marginal quality, a few α helices could be identified and manually docked. Further building of this model with Parrot (Zhang, Cowtan, & Main, 1997) and Buccaneer (Cowtan, 2006) resulted in the addition of a few β strands but with minimal registry of the primary sequence. This partial model could be successfully used to find a molecular replacement solution for the 1.75 Å BamL data set. Automated and manual rebuilding yielded a nearly complete model. The crystallographic phases of BpumL and other protein-ligand complexes were determined by the molecular replacement method (McCoy, Grosse-Kunstleve, Adams, Winn, Storoni, & Read, 2007) in the Phenix suite (Adams, et al., 2011), using BamL as searching model. Rounds of manual rebuilding using COOT (Emsley & Cowtan, 2004) were interspersed with crystallographic refinement using REFMAC5 (Murshudov, et al., 2011) until the free R factor dropped below 0.30. The validity of

all models was routinely determined with PROCHECK (Laskowski, Rullmannn, MacArthur, Kaptein, & Thornton, 1996), MOLPROBITY (Chen, et al., 2010) and by using the free R factor to monitor improvements during building and crystallographic refinement.

Kinetic Analysis of BamL and BpumL Methyltransferases

Kinetic analysis of BamL and BpumL utilized an enzyme-coupled continuous spectrophotometric assay (Dorgan, et al., 2006). Reaction mixtures typically consisted of 50 mM Tris pH 7.8, 200 μM $MnSO_4$, 5 μM *E. coli* pfs nucleosidase, 5 μM *B. subtilis* adenine deaminase, 80 μM SAM, and varying concentrations of the Arg-Az_1, Arg-Az_3, and Arg-Az_5. The reaction progress was monitored at 37 °C and utilized 0.5 μM of either wild-type MBP-BamL or BpumL. Concentrations for the mutant MBP-BamL were as follows: 1.4 μM R42A, 1.325 μM T38F, 1.34 μM L132F, 1.36 μM L162F, and 1.8 μM Y182F. Reaction progress was monitored as changes in absorption at a wavelength of 265 nm. Initial velocity data were calculated and fitted to the Michaelis-Menten equation using the Origin software package (OriginLab). Due to absorption by the substrates at 265 nm, reaction using high concentrations (>80 μM) of substrates failed to yield reliable data.

Purification of Coupling Enzymes for Kinetic Assay

The *E. coli* pfs nucleosidase plasmid pQE60-pfs was a gift from Prof. J. Cronan. It contained a T5 promoter and DH5α strain was used for protein expression. The culture first grew at 37 °C in LB medium supplemented with 100 μg/mL ampcillin until $O.D._{600}$ reached 0.5. Overexpression of pfs was induced with 1 mM IPTG at 37 °C for 4 h. The cell harvest and disruption and protein purification were carried out in a manner similar to BpumL. The clean

fractions from affinity chromatography were pooled and dialyzed to storage buffer (20 mM Tris pH 7.5, 300 mM NaCl, 30% glycerol) overnight. The resulting enzyme was aliquoted, flash-frozen and stored in -80 °C.

The adenine deaminase gene from *Bacillus subtilis* was amplified and cloned into pET-MBP, modified from commercial pET-28b (Novagen). The resulting recombinant deaminase had a His-tag and MBP-tag fused at its N-terminus. The heterologous expression, cell disruption and protein purification procedures were the same as BpumL. The clean fractions from affinity chromatography were pooled and dialyzed to storage buffer (20 mM Tris pH 7.5, 300 mM NaCl, 30% glycerol) overnight. The resulting enzyme was aliquoted, flash-frozen and stored in -80 °C. The mass concentration of pfs and MBP-adenine deaminase were determined using Bradford protein assay and then converted to molar concentration.

Results

Structural Characterization of BamL and BpumL

In order to investigate the unexpectedly high substrate selectivity, we determined the crystal structures of BamL with a resolution of 1.75 Å and BpumL with a resolution of 1.9 Å, each in complex with SAH. Both of the structures contain a typical Rossmann fold at the core similar to the architecture of other methyltransferases (Fig. 3.2 A, B). A seven-stranded β sheet is in the center surrounded by six α helices. BamL and BpumL share 48% primary sequence identity and their structures are also superimposable with a RMSD of 1.1 Å. Structural homology search of BamL against PDB identified the closest structure as the bacterial Hen1 methyltransferase domain (PDB 3JWG) with a RMSD of 2.5 Å over 174 aligned Cα and the hypothetical bacterial YecO protein (PDB 1IM8) with a RMSD of 2.9 Å over 194 aligned Cα. The structural homology

mainly results from the conserved Rossmann fold in methyltransferases despite the marginal similarity across the primary sequence.

Unlike the architectures of typical small molecule methyltransferases, BamL and BpumL contain only the minimal characteristic structural features as methyltransferases without extra fold or domain. One molecule of SAH binds inside BamL in a pocket defined by S92 and F137 along the sides and a loop S112–A114 across the top (Fig. 3.2 C). R42 and H131 coordinate the carboxylate group of the homocysteine moiety through hydrogen bonds. D91 provides additional interaction with the 2' hydroxyl group of the ribose. N113 is in hydrogen-bonding distance with the amine group of the adenine ring. These residues are conserved in the plantazolicin *N*-methyltransferases (Fig. 3.3), mutations of which significantly compromise the methyltransferase activity. Interestingly, the trajectory of the homocysteine moiety of SAH is not collinear with the adenine ring but curves to the interior of BamL due to the F21 and loop G68–Q71, which is the conserved GxG loop in SAH-binding site (Kozbial & Mushegian, 2005). Consequently, the electrophilic methyl group of SAM is positioned at the junction of the SAM-binding pocket and a long tunnel that runs to the surface of the protein. The wall of this tunnel is lined up with a series of hydrophobic residues including F21, Y33, T38, L132, L162, Y182, L183, Y187 and L258 (Fig. 3.2 E). These residues are highly conserved between BamL and BpumL, except that the L183 is replaced by I181 in BpumL. Most of them are also retained in other annotated PZN *N*-methyltransferases, indicating the existence and importance of the hydrophobic tunnel (Fig. 3.3). Since the depth of the tunnel is sufficient for accommodating part of the desmethylPZN, it is proposed to be the substrate-binding tunnel. The pinch point of the tunnel is roughly 4 Å and barely fit any residues (Fig. 3.2 F). It greatly limits the substrate spectrum to peptides containing planar azole structures following the N-terminal Arg and explains why Ala, Cys or Pro

substitutions abolish the methyltransferase activity. The amine group of the Arg is thus believed to situated at the junction of this tunnel and the SAM-binding pocket together with the methyl group of SAM. Supporting this postulate, a shorter, perpendicular cavity is found here to fit the side chain of Arg.

Kinetic Analysis of Truncated DesmethylPZN Analogs

Initial attempt on kinetic study was hindered by the poor aqueous solubility of desmethylPZN until the better-behaved truncated desmethylPZN analogs were successfully synthesized chemically and shown high affinity to the enzyme (Sharma, Blair, & Mitchell, 2013). Together with the hypothesis that the substrate-binding tunnel only accommodates part of the desmethylPZN, we conducted kinetic analysis with each of the truncated substrate analogs Arg-Az_1, Arg-Az_3 and Arg-Az_5 using an enzyme coupled assay to continuously monitor SAH production (Fig. 3.4). The coupling enzymes *E. coli* SAH nucleosidase and *B. subtilis* adenine deaminase have a k_{cat} of 35.2 ± 0.92 s^{-1} and 4.12 ± 0.10 s^{-1} respectively (Dorgan, et al., 2006), greater than that of the two enzymes assayed. Under the assay condition, BamL showed a catalytic efficiency (k_{cat}/K_M) of 3.46 × 10^3 M^{-1} s^{-1} for Arg-Az_1, 5.75 × 10^3 M^{-1} s^{-1} for Arg-Az_3 and 7.53 × 10^3 M^{-1} s^{-1} for Arg-Az_5, increasing with the number of heterocycles (Table 3.1). While the k_{cat} for the three substrates are comparable, K_M value decreases with the increasing length of the substrate, indicating tighter interactions between the enzyme and the longer substrate. BpumL was only assayed with Arg-Az_3, displaying a catalytic efficiency of 6.38 × 10^3 M^{-1} s^{-1} comparable to that of BamL with the same substrate.

Co-crystal Structures of PZN Methyltransferases with Truncated Substrates

Based on the enzyme activity and high affinity to truncated desmethylPZN substrates, we conducted co-crystallization study of both BamL and BpumL with the three assayed substrates to gain structural insights for the strict substrate selectivity. To obtain substrate-bound ternary complex, BamL was incubated with SAH and Arg-Az_1 or Arg-Az_3 and BpumL with SAH and Arg-Az_1 prior to co-crystallization. To obtain product-bound ternary complex, equal concentrations of SAM and Arg-Az_3 or Arg-Az_5 were incubated with BpumL according to the pseudo single-turnover assay of BamL with desmethylPZN, so that Me-Arg-Az_3 or Me-Arg-Az_5 is expected in the structure together with SAH. The structures of BamL and BpumL in complex with SAH and either of the Arg-Az_1 and Arg-Az_3 were determined with a resolution between 1.5 Å and 1.8 Å. The structure of BpumL with SAH and Arg-Az_5 was obtained at 1.75 Å resolution. In the ternary structures, clear and continuous electron density was observed for the N-terminal Arg and all of the heterocycles of Arg-Az_1 and Arg-Az_3 in the immediate vicinity of SAH; yet only the first four out of the five heterocycles of Arg-Az_5 in BpumL-Arg-Az_5-SAH structure showed strong electron density with the C-terminal fifth ring displaying much weaker electron density.

The structures of the ternary complexes didn't show huge conformational changes in comparison with the BamL-SAH or BpumL-SAH discussed previously (Fig. 3.5 A, B; Fig. 3.6 A-C). The architecture of the Rossmann fold shapes two main cavities: one cavity is the SAM/SAH-binding pocket that accommodates SAH in a bent pose; the second cavity, connecting the first cavity, is the polyazolic substrate-binding tunnel as predicted (Fig. 3.5 C, Fig. 3.6 D). This hydrophobic tunnel is formed by the capping helices of the Rossmann fold and is perpendicular to the central β sheet. The (poly)azolic substrate binds in the tunnel in an extended manner with the α-amine of N-terminal Arg positioned at the base of the tunnel. The

location of the α-amine is at the junction of the tunnel and the pocket and in proximity to the electrophilic methyl group of SAM, expected from the orientation of the SAH by-product. Two polar residues D34 and Y182 (BamL numbering) are in hydrogen-bonding distance with it and serve as potential active site bases for proton abstraction (Fig. 3.7 A, C; Fig. 3.8 A, C, E). The perpendicular cavity observed previously at the junction holds the side chain of Arg in a planar conformation through a combination of van der Waals interactions and hydrogen-bonding network. BamL residues L162 and Y187 establish one side of the cavity and a short 3_{10} helix spanning F13–F137 defines the opposite boundary. The guanidinum group of Arg is in hydrogen-bonding interactions with the side chains of E161, Q186 and S190, and the main chain carbonyl group of Q186 (Fig. 3.7 B, D; Fig. 3.8 B, D, F). The dimensions of the cavity and its extensive interactions with the side chain of the N-terminal residue support the lack of BamL methyltransferase activity towards polyazolic substrates with Lys, ornithine or homoarginine substitution at this position.

The polyazolic substrate-binding tunnel is established by a group of hydrophobic residues and backed by a long, bent α helix (Fig. 3.7 A-D, Fig. 3.8 A-F). The thiazole at position 2 and methyloxazole at position 3 of the substrate are surrounded by hydrophobic residues including Y33, L132, I141, V160, L162, Y182, L183 and Y187. The methyloxazole is also in van der Waals contacts with L132 and L258. There is not much interaction between the tunnel and the rest of the substrate, except that thiazole at position 4 is in van der Waals contact with L258 and methyloxazole at position 5 forms aromatic-aromatic interaction with Y247 of BpumL, which is not a conserved residue in BamL. The depth of the tunnel could well accommodate 4 azole rings with the fifth ring at the edge, which partially explains its weak electron density in the structure (Fig. 3.6 D). Predicted from the enzyme-substrate interactions, the substitution of either of the

two thiazoles by methyloxazole would lead to steric clashes and thus truncated desmethylPZN with Thr at position 2 or 4 won't be favorable methyltransferase substrate.

The attempt on crystallizing product-bound ternary complexes of BpumL under single turnover condition was successful and monomethylated product Me-Arg-Az_3 or Me-Arg-Az_5 was clearly observed with SAH in the active site, with a nearly identical orientation with the corresponding desmethyl substrate (Fig. 3.8 C-F). The monomethylated α-amine group of Arg is situated in a relatively hydrophobic environment created by BpumL residues F20, L130 and L134. The dimensions of the pocket are sufficient to fit the dimethylation on Arg of the substrate. Two polar residues D33 and Y180 are in hydrogen-bonding distance with the α-amine as in the substrate-bound structures.

Conformational Change of a Capping Loop upon Substrate Binding

In BamL, the substrate-binding tunnel is covered by a flexible region of the enzyme spanning residues T244–A257, over the position of the third azole ring through the fifth azole ring. It is close to the C-terminus of the enzyme and links an α helix to the Rossmann fold core. In the absence of polyazolic substrate, this region is a flexible loop with no obvious interactions with other parts of the enzyme (Fig. 3.9 A). With the presence of Arg-Az_1, residues G247–M255 change to an α helix orthogonal to the helix it links, to better encapsulate the substrate (Fig. 3.9 B). With the presence of Arg-Az_3, only residues S245–G248 form a one-turn helix joining the helix it links with the remainder as a loop, since the lengthy substrate breaks the trajectory of the helix (Fig. 3.9 C). The result of this conformational change is the optimized placement of M255 to facilitate the packing of C-terminal azole and the establishment of an appropriate hydrophobic pocket according to the length of the substrate. This polymorphic region is among the least

conserved regions in primary sequence alignment of the PZN methyltransferases with several missing the entire C-terminal extension (Fig. 3.3).

Kinetic Analysis of Active Site Variants with Substrate Arg-Az_3

The ternary structures reported here reveal several residues potentially important for catalysis or substrate binding in the vicinity of the bound polyazolic substrate. Based on the structure, several site-specific BamL variants were generated for kinetic analysis with Arg-Az_3 as substrate (Table 3.2). T38 is located in the aforementioned long, bent helix that supports the substrate-binding tunnel and is in van der Waals contact with the first two azoles of the substrate. T38F mutant shows a 17-fold decrease in catalytic efficiency. Likewise, Phe substitution of L132 or L162 that are in contact with the first two azoles also leads to decreased catalytic efficiency by 16 fold and 48 fold, respectively. In these cases, the replacement with bulkier Phe side chain potentially narrows the substrate-binding tunnel and creates steric clashes to restrict the access of substrate to the methylation site. Y182 is in hydrogen-bonding distance with the α-amine of the substrate and Y182F mutation results in a 17-fold decrease in catalytic efficiency, consistent with its potential role as an active site base for α-amine deprotonation. Y182 is not conserved in some annotated PZN methyltransferases. A Phe appears in this position in two of the transferases, but neighboring Tyr residues may play a similar role to Y182. Another residue R42 in coordinating the carboxyl group of SAH/SAM was also investigated by Ala substitution. Mutation R42A results in a decreased catalytic efficiency by 12 fold, indicating its importance in catalysis.

Discussion

Plantazolicin is a linear azol(in)e-containing peptide produced by the plant growth-promoting rhizobacterium *Bacillus amyloliquefaciens* FZB42 (Scholz, et al., 2011). It is a narrow-spectrum antibiotic potently against the anthrax-causing pathogen *Bacillus anthracis*. The bioactivity of plantazolicin is heavily dependent on the dimethylation at the α-amine of the N-terminal Arg (Molohon, et al., 2011). The dimethylation is catalyzed by the *N*-methyltransferase BamL in the PZN gene cluster in a leader peptide independent manner, which requires either the precise recognition of substrate or broad substrate tolerance by the enzyme. The latter strategy is adopted by a previously reported linaridin *N*-methyltransferase CypM (Zhang & van der Donk, 2012), which dimethylates the α-amine of the N-terminal Ala in cypemycin biosynthesis. In contrast, biochemical studies showed that BamL is highly selective towards its cognate substrate, Arg with succeeding polyazole. Substitution of the azolic structure with any unmodified residue completely abolished the methyltransferase activity. Interestingly, BamL retains methyltransferase activity towards truncated desmethylPZN analogs, Arg-Az_5, Arg-Az_3, and Arg-Az_1. It even catalyzes methylation of Arg amide. The kinetic studies presented here demonstrated that the truncated desmethylPZN analogs are tightly bound substrates with low-micromolar affinity and are effectively turned over by BamL and BpumL with comparable k_{cat}. The K_M showed a decreasing trend with increasing numbers of azole rings, indicating the possible involvement of all five azoles in enzyme-substrate interactions. Although Arg and the first azole ring already guarantee the substrate binding, the presence of the rest four azoles doubles the substrate affinity. To gain insights in the structural basis of this unusual substrate specificity, the co-crystallization of BamL in complex with SAH and truncated desmethylPZN analogs were attempted and reported here, despite the failure in co-crystallization with the native substrate desmethylPZN.

The BamL overall structure, as well as the structure of its homolog BpumL, appears as a minimal Rossmann fold typical in SAM-dependent methyltransferases. Unlike most small molecule methyltransferases, BamL contains no elaborated auxiliary folds for substrate recognition. SAH is coordinated mainly through hydrogen bonds in the binding pocket, with the homocysteine moiety bent towards the interior of BamL. The structure feature that determines the substrate selectivity is the deep, narrow, hydrophobic substrate-binding tunnel formed by the capping helices of the Rossman fold. The tunnel is perpendicular to the central β sheet and runs from the surface to the interior of the protein. It connects the SAH/SAM-binding pocket at the base and an additional narrow cavity protrudes perpendicularly at the joint.

In the co-crystal structure of BamL/BpumL with the truncated substrate analogs, the planar azolic substrate is tightly situtated in the aforementioned hydrophobic tunnel, with the side chain of N-terminal Arg fits in the perpendicular narrow cavity. The effective substrate binding is made through the extensive hydrophobic interactions with the polyazolic structure and the hydrogen-bonding interactions with Arg side chain. Fist, the dimension of the substrate-binding tunnel substantially restricts the access to catalytic site by peptides with unmodified residues. The steric restrains created by the residues lining the tunnel would energetically disfavor the accommodation of these substrates unless considerable conformational changes occur in the enzyme. Also, as indicated by the kinetic analysis of BamL mutants, the dimensions of the tunnel are precisely designed for the restrained planar polyazolic substrate. The native substrate would also be blocked if the tunnel is further narrowed by introduction of bulkier residues on the wall. Second, the residues lining the substrate-binding tunnel are mainly hydrophobic, which prefers the placement of hydrophobic substrate in the tunnel. Any charges or polar groups would be unfavorable. This provides an additional explanation for the selectivity towards polyazolic

substrate, because heterocyclized residues are apparently less hydrophilic compared with their unmodified counterparts. It also rationalizes the result that Arg thiazole with a terminal amide is the minimal BamL substrate (Piwowarska, Banala, Overkleeft, & Süssmuth, 2013) and that Arg amide is only turned over under single-turnover condition. In contrast, the free amino acid Arg, bearing a carboxylate group, is not accepted as a substrate, despite the satisfaction of steric requirement. Last, the side chain of N-terminal Arg fits tightly in the cavity perpendicular to the substrate tunnel. The guadinino group at the base of the cavity forms hydrogen bonds with D161 and S190, as well as the backbone carbonyl group of Q186. This extensive hydrogen-bonding network observed both in BamL and BpumL co-crystal structures justifies the strict requirement of Arg at N-terminus of the substrate. Substitution with either ornithine or Lys disrupts the contacts and results in poor substrate. Similarly, homoArg substitution results in an extra methylene group in side chain that fails to meet the steric limitation of the cavity.

At the joint of the SAH/SAM-binding pocket and substrate-binding tunnel, the S atom of SAH and the α-amine are located in proximity to each other. BamL residues D34 and Y182 are in hydrogen-bonding distance to the α-amine and likely facilitate the methyltransferase reaction by deprotonating the α-amine. Mutation Y182 to Phe significantly impedes the catalytic efficiency of BamL. Interestingly, Phe naturally occurs in two BamL homologs CurL and BlinL, which indicates a similar role may be played by other residues in the vicinity, such as the Y21 compared to F21 in BamL. The catalytic cycle is thus proposed as following (Fig. 3.10). A base at active site deprotonates the α-amine of the bound substrate and facilitates its nucleophilic attack onto the methyl group of SAM. The resulting SAH is then replaced by a second molecule of SAM. The monomethylated α-amine is further deprotonated and methylated to the dimethylated α-amine, which is less favored in the hydrogen-bonding network. The increased

hydrophobicity of the product disfavors binding in the catalytic site and enables the release of the product.

In summary, the co-crystal structures of BamL/BpumL with the truncated desmethylPZN analogs elucidate the strict substrate selectivity of plantazolicin methyltransferase. PznL is designed with such structural features that it does not "cross-react" with other peptide substrates, nor disturb other cellular processes. It is specifically expressed for a single purpose and not exploited by other processes, ensuring efficient plantazolicin production. The structure information also provides clues for rational engineering of plantazolicin. The substrate-binding tunnel forms extensive interactions mainly with the first five or six residues of the substrate and the rest of the residues are exposed. Substituting these residues or modifying the length of the substrate would be a feasible method to generate plantazolicin library for bioactivity assessment.

Figures

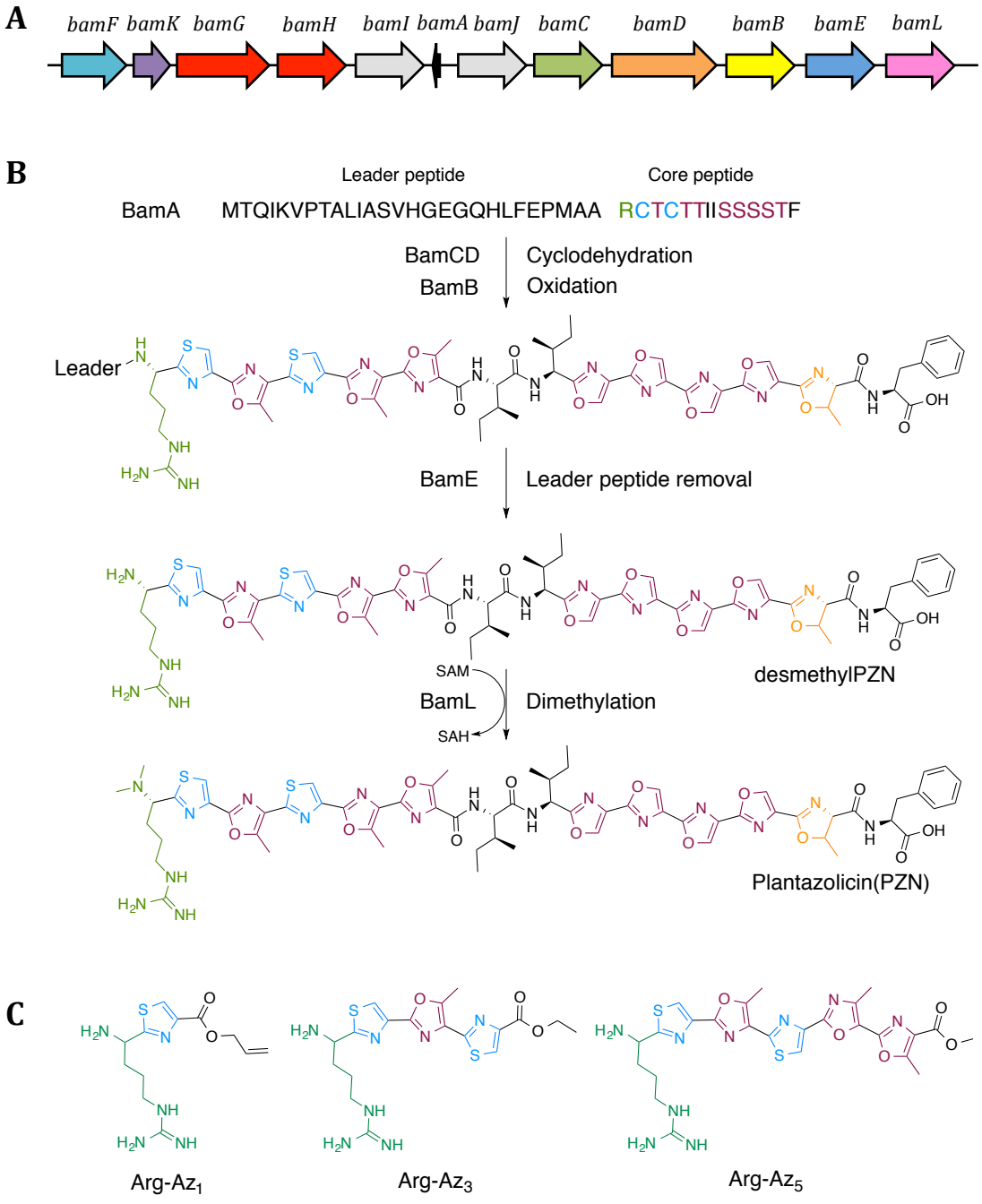

Fig. 3.1 **A.** Plantazolicin biosynthetic gene cluster from *Bacillus amyloliquefaciens* FZB42. *bamA* encodes the precursor peptide; *bamD* encodes the heterocyclase (cyclodehydratase) and

bamC encodes the docking protein; *bamB* encodes the FMN-dependent dehydrogenase; product of *bamE* is a zinc-dependent protease; *bamL* encodes the SAM-dependent methyltransferase; *bamF* is a immunity gene; *bamK* encodes a transcriptional regulator; *bamGH* encode ABC transporter. **B.** The general biosynthetic pathway of plantazolicin. Precursor peptide BamA undergoes cyclodehydration and oxidation by BamBCD complex to form two patches of poly azol(in)e. Removal of the leader peptide generates a N-terminal amine. It is dimethylated by BamL and the resulting PZN is in active form. **C.** Synthetic desmethylPZN analogs, Arg-Az_1, Arg-Az_3, and Arg-Az_5. They bind BamL with low micromolar affinity and can be converted by dimethylated products by BamL.

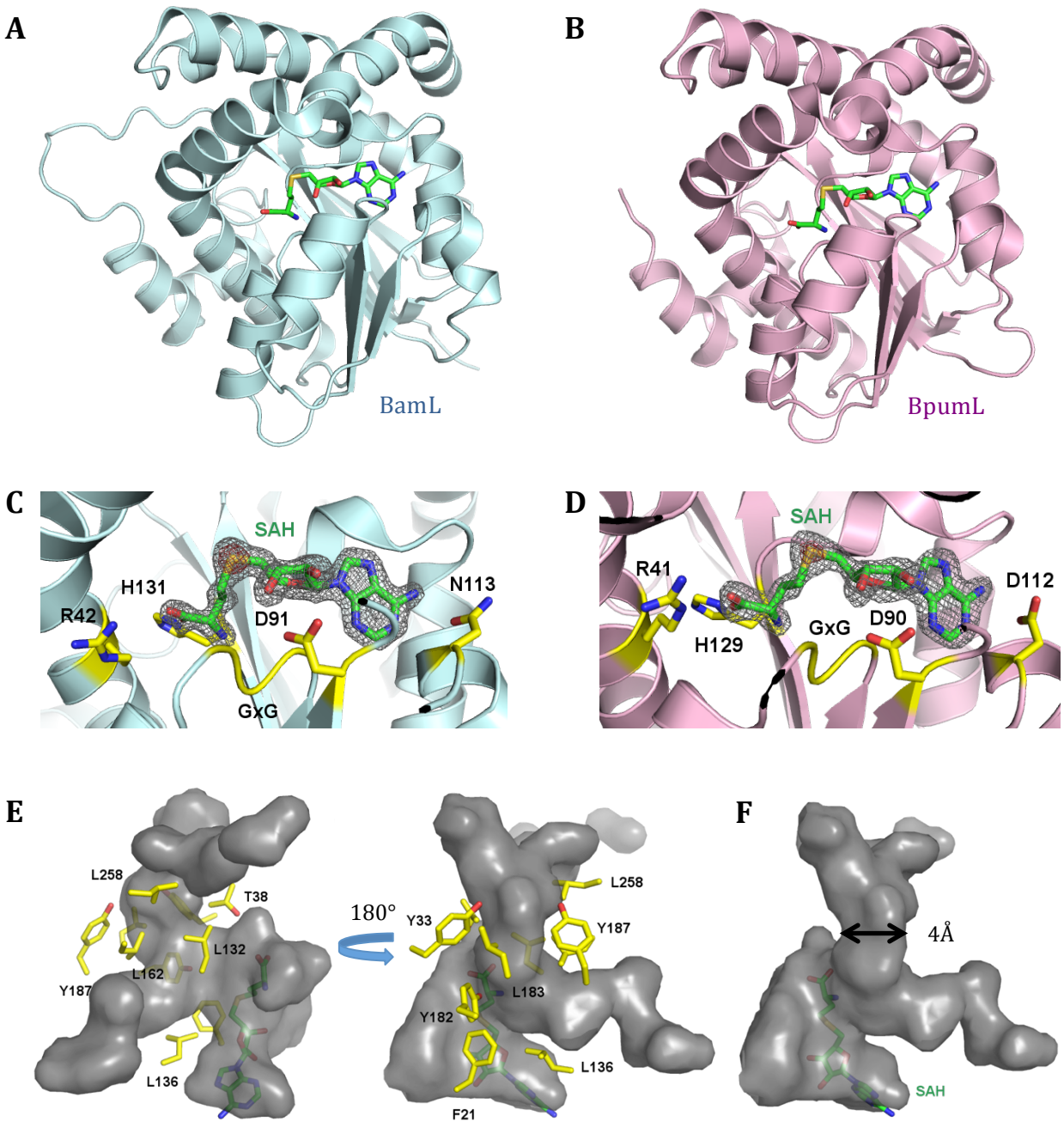

Fig. 3.2 **A.** Overall structure of BamL in complex with SAH. SAH, shown as sticks, is located in a pocket in the Rossmann fold. **B.** Overall structure of BpumL in complex with SAH. SAH, shown as sticks is located in a pocket in the Rossmann fold. **C.** Coordination of SAH in the binding pocket of BamL. The 2*Fo-Fc* map is shown with the stick representation of SAH (green).

The map contoured at 1.5σ is shown in gray and the map contoured at 4σ is shown in red. BamL residues R42 and H131 form hydrogen bonds with the carboxylate group of homocysteine of SAH. D91 interacts with the 2' hydroxyl group of the ribose moiety. N113 interacts with the adenine ring through hydrogen bonds. The conserved GxG motif in SAH/SAM-binding pocket of SAM-dependent methyltransferases is highlighed in yellow. **D.** Coordination of SAH in the binding pocket of BpumL. The 2*Fo-Fc* map is shown with the stick representation of SAH (green). The map contoured at 1.5σ is shown in gray and the map contoured at 4σ is shown in red. BpumL residues R41 and H129 form hydrogen bonds with the carboxylate group of homocysteine of SAH. D90 interacts with the 2' hydroxyl group of the ribose moiety. D112 interacts with the adenine ring through hydrogen bonds. The conserved GxG motif in SAH/SAM-binding pocket of SAM-dependent methyltransferases is highlighed in yellow. **E.** A tunnel in BamL lined with hydrophobic residues is connected to the SAH-binding pocket and the joint is at the position of the methyl group of SAM. Another cavity protrudes from the joint, perpendicular to the hydrophobic tunnel. **F.** The hydrophobic tunnel runs to the surface of the BamL and is relatively narrow. The dimension of the pinch point is about 4 Å.

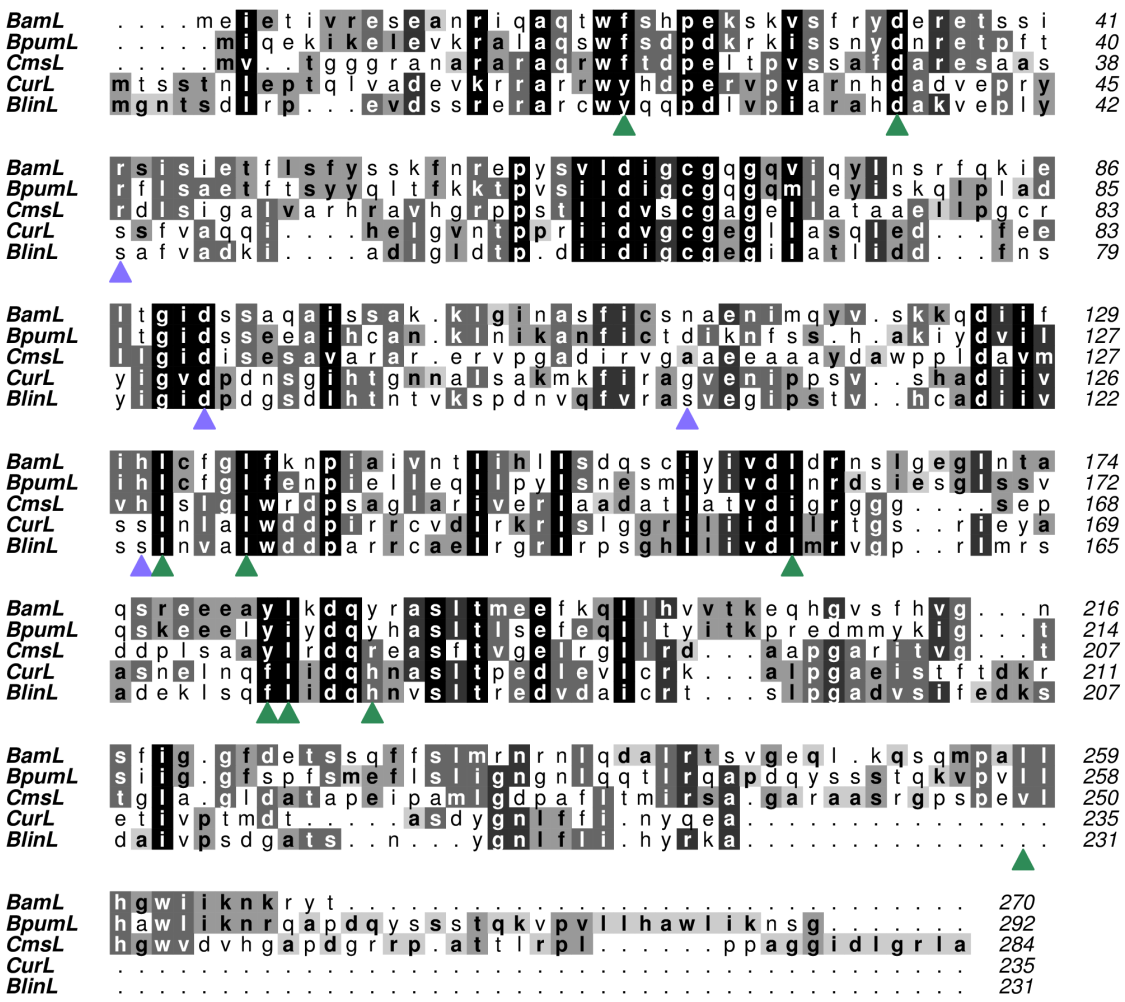

Fig. 3.3 Sequence alignment of the annotated plantazolicin methyltransferases PznLs from different bacteria. BamL is from *Bacillus amyloliquefaciens* FZB42; BpumL is from *Bacillus pumilus* ATCC7061; CmsL is from *Clavibacter michiganensis* subsp. *sepedonicus*; CurL is from *Corynebacterium urealyticum* DSM 7109; BlinL is from *Brevibacterium linens* BL2. Residues important for SAH/SAM binding are indicated with purple arrowheads. Residues involved in polyazolic substrate coordination are indicated with green arrowheads.

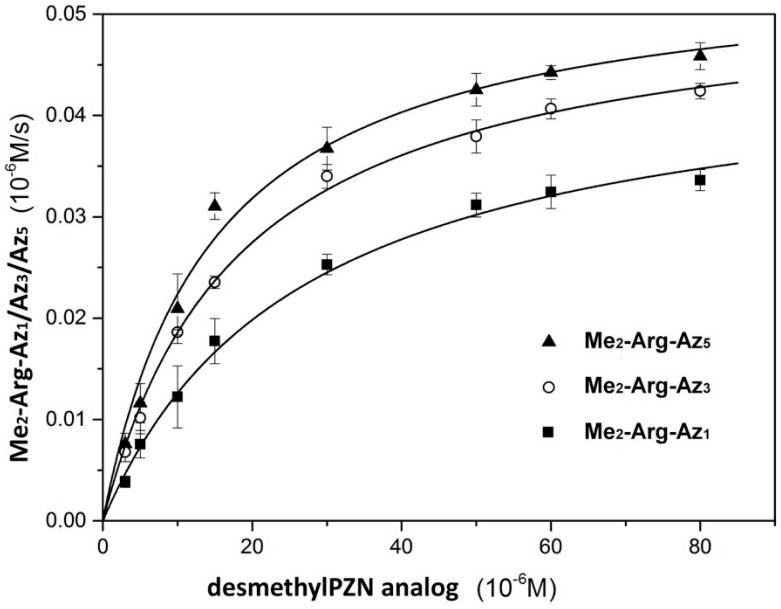

Fig. 3.4 Michaelis-Menten curves of BamL assayed with truncated PZN analogs as substrates. The co-substrate SAM is provided at 80 μM, which is 10 × K_d calculated from the isothermal titration calorimetry measurement. The coupling enzyme pfs nucleosidase cleaves the SAH product into adenine and *S*-ribosyl homocysteine. The former one is further converted to hypoxanthine by adenine deaminase, coupled with a decrease of absorption at 265 nm. This correlation between absorption change and SAH generation allows for the calculation of SAH produced in the reaction, which is then correlated to the production of dimethylated PZN analogs.

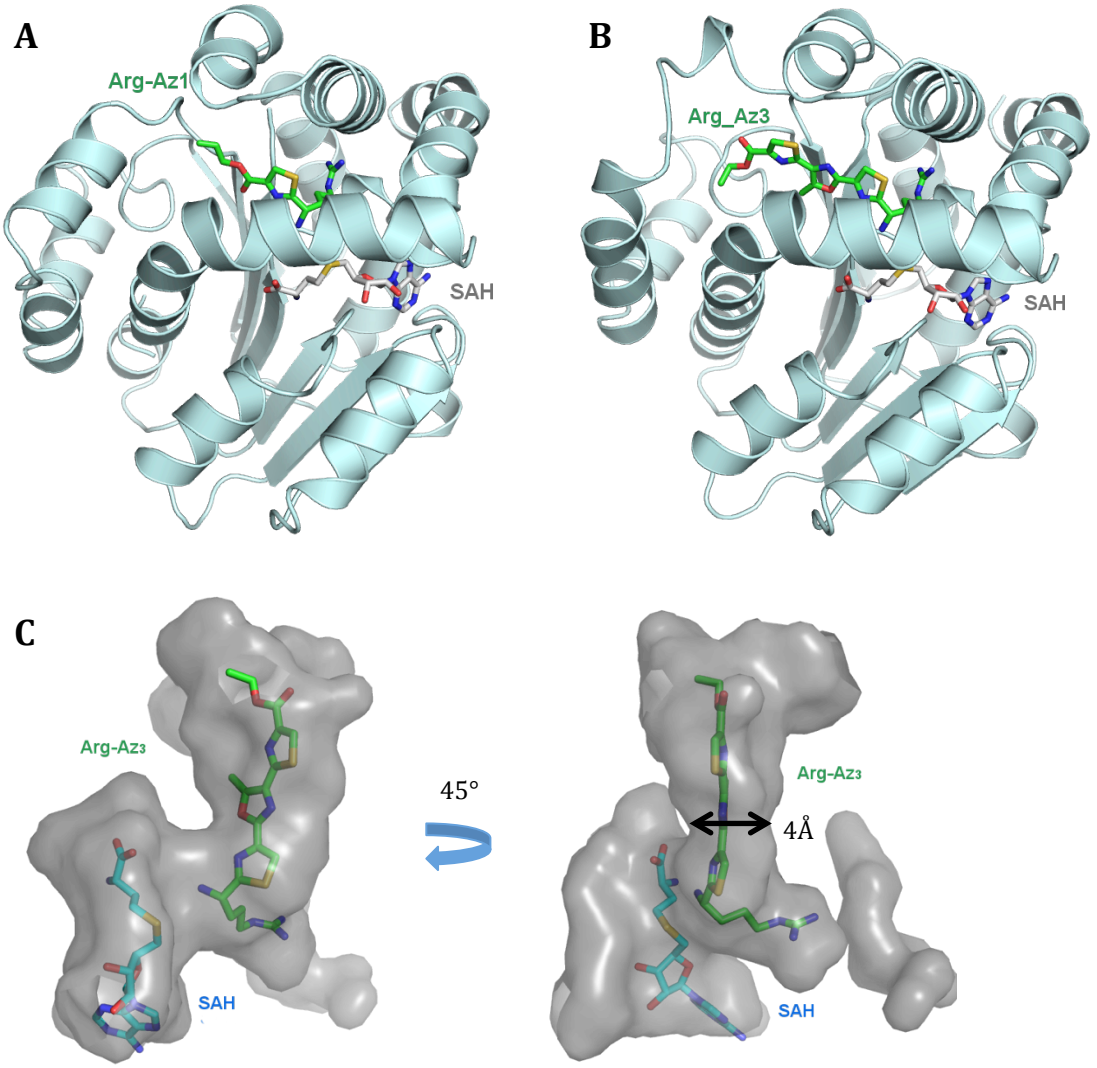

Fig. 3.5 **A.** Co-crystal structure of BamL with SAH (gray sticks) and Arg-Az_1 (green sticks). **B.** Co-crystal structure of BamL with SAH (gray sticks) and Arg-Az_3 (green sticks). **C.** SAH (cyan) and Arg-Az_3 (green) shown in the BamL SAH/SAM-binding pocket and substrate-binding tunnel, respectively. The α-amine of the N-terminal Arg is in proximity to the S of SAH, and potentially the methyl group of SAM. The Arg side chain is accommodated in the perpendicular cavity. The planar polyazolic substrate fits properly in the narrow tunnel.

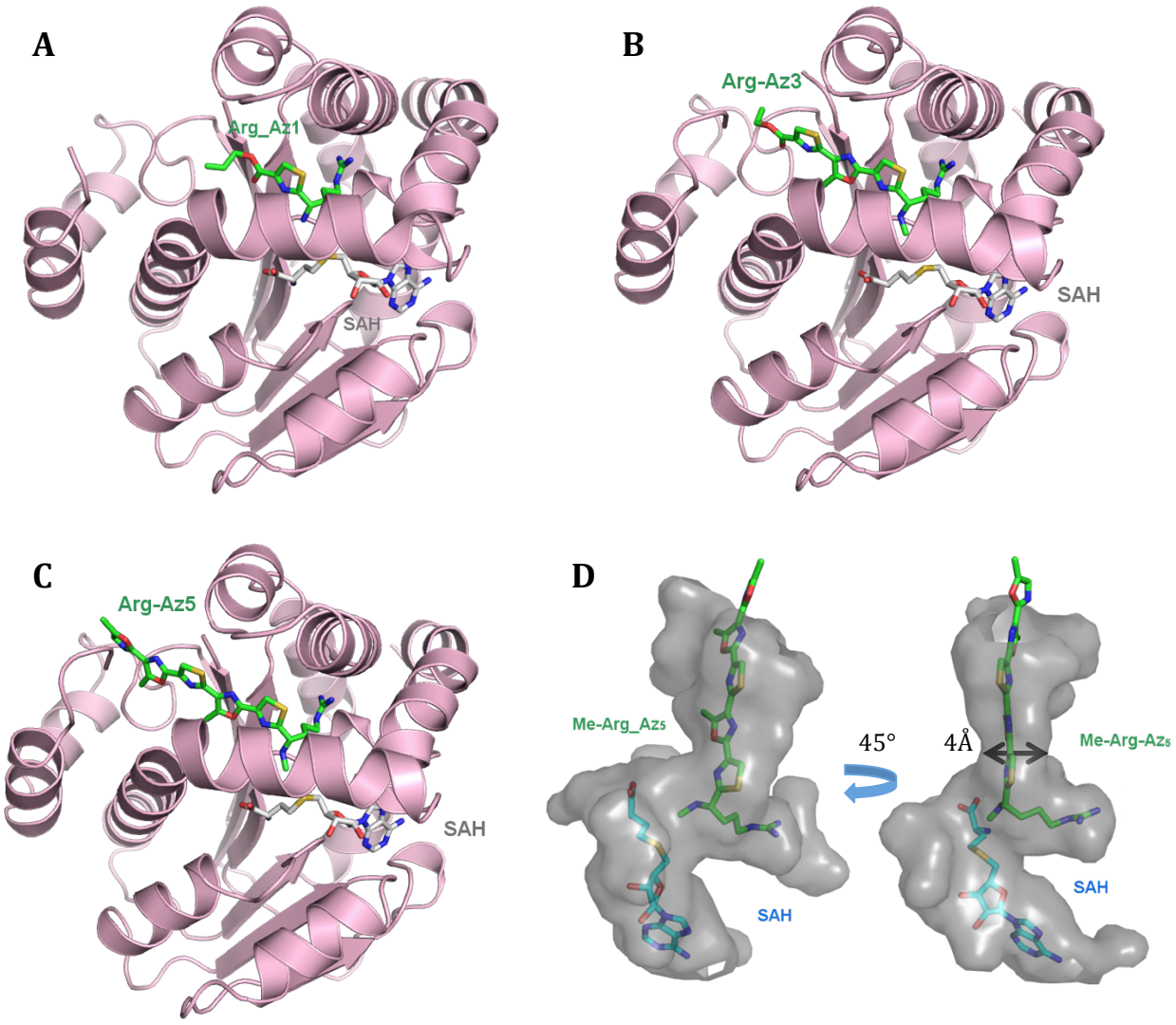

Fig. 3.6 **A.** Co-crystal structure of BpumL with SAH (gray sticks) and Arg-Az_1 (green sticks). **B.** Co-crystal structure of BpumL with SAH (gray sticks) and Me-Arg-Az_3 (green sticks). **C.** Co-crystal structure of BpumL with SAH (gray sticks) and Me-Arg-Az_5 (green sticks). **D.** SAH (cyan) and Me-Arg-Az_5 (green) shown in the BpumL SAH/SAM-binding pocket and substrate-binding tunnel, respectively. The monomethylated α-amine of the N-terminal Arg is in proximity to the S of SAH, and potentially the methyl group of SAM. The Arg side chain is accommodated in the perpendicular cavity. The planar polyazolic substrate fits properly in the narrow

hydrophobic tunnel. The depth of the tunnel fits the first four azole rings and thus the fifth ring is exposed to solvent, making almost no contact with the enzyme. Based on the co-crystal structure, only the first polyazole patch interacts with the enzyme and the following residues are fully exposed to solvent. The flexibility of the second half of desmethylPZN is one possible reason for the failure of co-crystallization with full-length desmethylPZN.

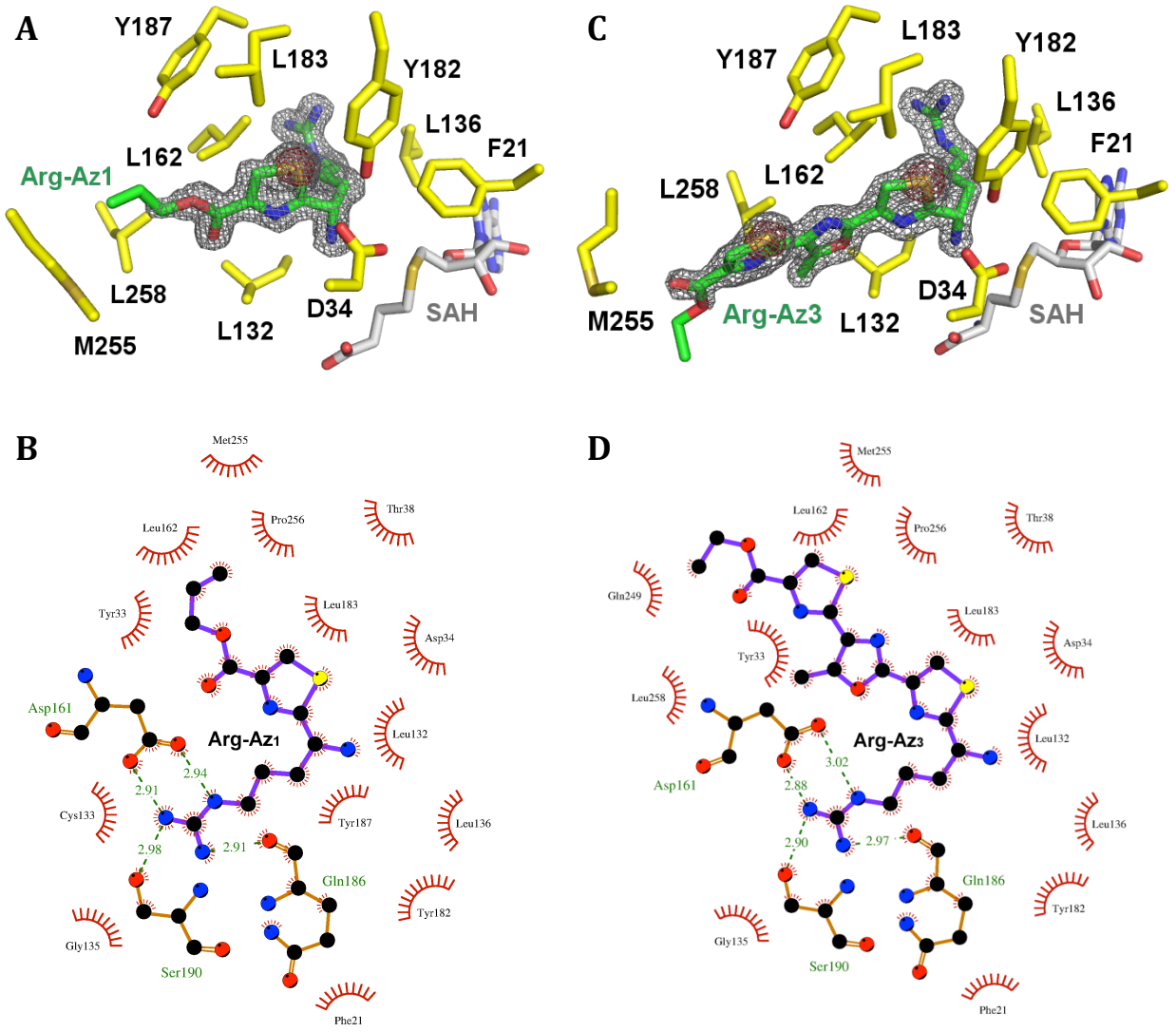

Fig. 3.7 **A.** Coordination of Arg-Az$_1$ (green sticks) in the substrate-binding tunnel of BamL. The 2*Fo-Fc* map of Arg-Az$_1$ contoured at 1.5σ is shown in gray and the map contoured at 4σ is shown in red. Arg-Az$_1$ is coordinated by a cluster of hydrophobic residues in the tunnel. SAH is shown as gray sticks. **B.** LIGPLOT diagram (Laskowski & Swindells, 2011) showing the interactions between BamL and Arg-Az$_1$ (purple). Carbon, oxygen, nitrogen and sulfur atoms are colored in black, red, blue and yellow, respectively. Hydrogen bonds are shown in dashed green lines with indicated bond lengths. Hydrophobic interactions are shown as red arcs with radiating

spikes. **C.** Coordination of Arg-Az_3 (green sticks) in the substrate-binding tunnel of BamL. The 2*Fo-Fc* map of Arg-Az_3 contoured at 1.5σ is shown in gray and the map contoured at 4σ is shown in red. Arg-Az_3 is coordinated by a cluster of hydrophobic residues in the tunnel. SAH is shown as gray sticks. **D.** LIGPLOT diagram (Laskowski & Swindells, 2011) showing the interactions between BamL and Arg-Az_3 (purple). Carbon, oxygen, nitrogen and sulfur atoms are colored in black, red, blue and yellow, respectively. Hydrogen bonds are shown in dashed green lines with indicated bond lengths. Hydrophobic interactions are shown as red arcs with radiating spikes.

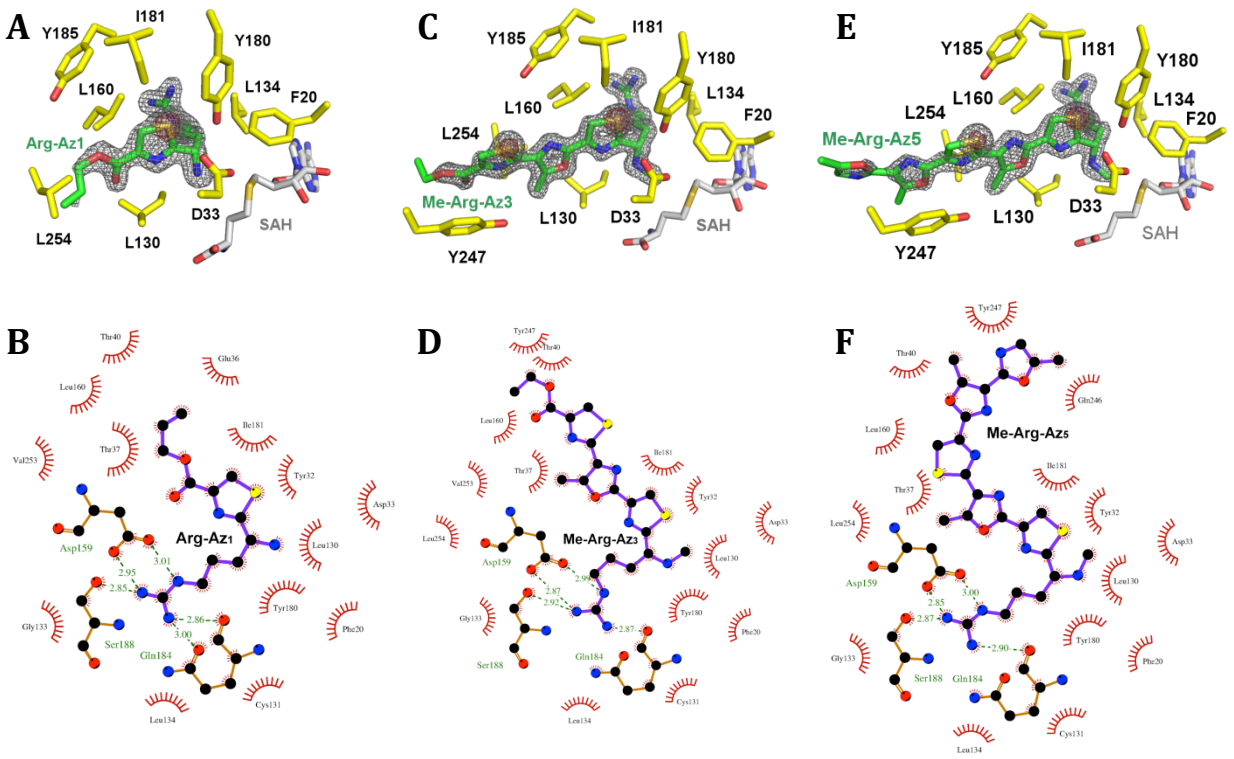

Fig. 3.8 **A.** Coordination of Arg-Az_1 (green sticks) in the substrate-binding tunnel of BpumL. The 2*Fo-Fc* map of Arg-Az_1 contoured at 1.5σ is shown in gray and the map contoured at 4σ is shown in red. Arg-Az_1 is coordinated by a cluster of hydrophobic residues in the tunnel. SAH is shown as gray sticks. **B.** LIGPLOT diagram (Laskowski & Swindells, 2011) showing the interactions between BpumL and Arg-Az_1 (purple). Carbon, oxygen, nitrogen and sulfur atoms are colored in black, red, blue and yellow, respectively. Hydrogen bonds are shown in dashed green lines with indicated bond lengths. Hydrophobic interactions are shown as red arcs with radiating spikes. **C.** Coordination of Me-Arg-Az_3 (green sticks) in the substrate-binding tunnel of BpumL. The 2*Fo-Fc* map of Me-Arg-Az_3 contoured at 1.5σ is shown in gray and the map contoured at 4σ is shown in red. Me-Arg-Az_3 is coordinated by a cluster of hydrophobic residues in the tunnel. SAH is shown as gray sticks. **D.** LIGPLOT diagram (Laskowski & Swindells, 2011) showing the interactions between BpumL and Me-Arg-Az_3 (purple). Carbon, oxygen, nitrogen

and sulfur atoms are colored in black, red, blue and yellow, respectively. Hydrogen bonds are shown in dashed green lines with indicated bond lengths. Hydrophobic interactions are shown as red arcs with radiating spikes. **E.** Coordination of Me-Arg-Az_5 (green sticks) in the substrate-binding tunnel of BpumL. The 2*Fo-Fc* map of Me-Arg-Az_5 contoured at 1.5σ is shown in gray and the map contoured at 4σ is shown in red. Me-Arg-Az_5 is coordinated by a cluster of hydrophobic residues in the tunnel. SAH is shown as gray sticks. **F.** LIGPLOT diagram (Laskowski & Swindells, 2011) showing the interactions between BpumL and Me-Arg-Az_5 (purple). Carbon, oxygen, nitrogen and sulfur atoms are colored in black, red, blue and yellow, respectively. Hydrogen bonds are shown in dashed green lines with indicated bond lengths. Hydrophobic interactions are shown as red arcs with radiating spikes.

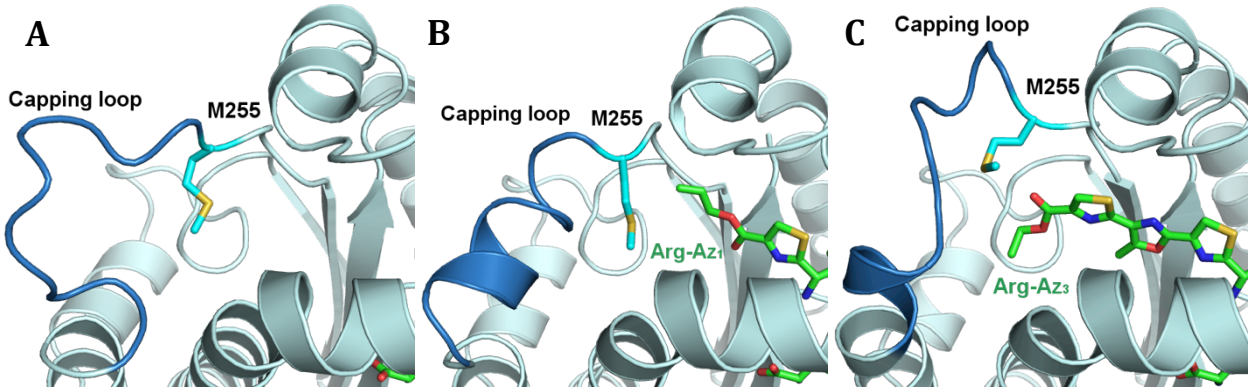

Fig. 3.9 **A.** A polymorphic region (blue) in BamL forms a loop and caps the hydrophobic substrate-binding tunnel in the absence of substrate. **B.** The polymorphic region (blue) in BamL forms an α helix in the presence of Arg-Az_1. **C.** The polymorphic region (blue) in BamL forms a one-turn α helix in the presence of Arg-Az_3 with the movement M265 to accommodate the substrate.

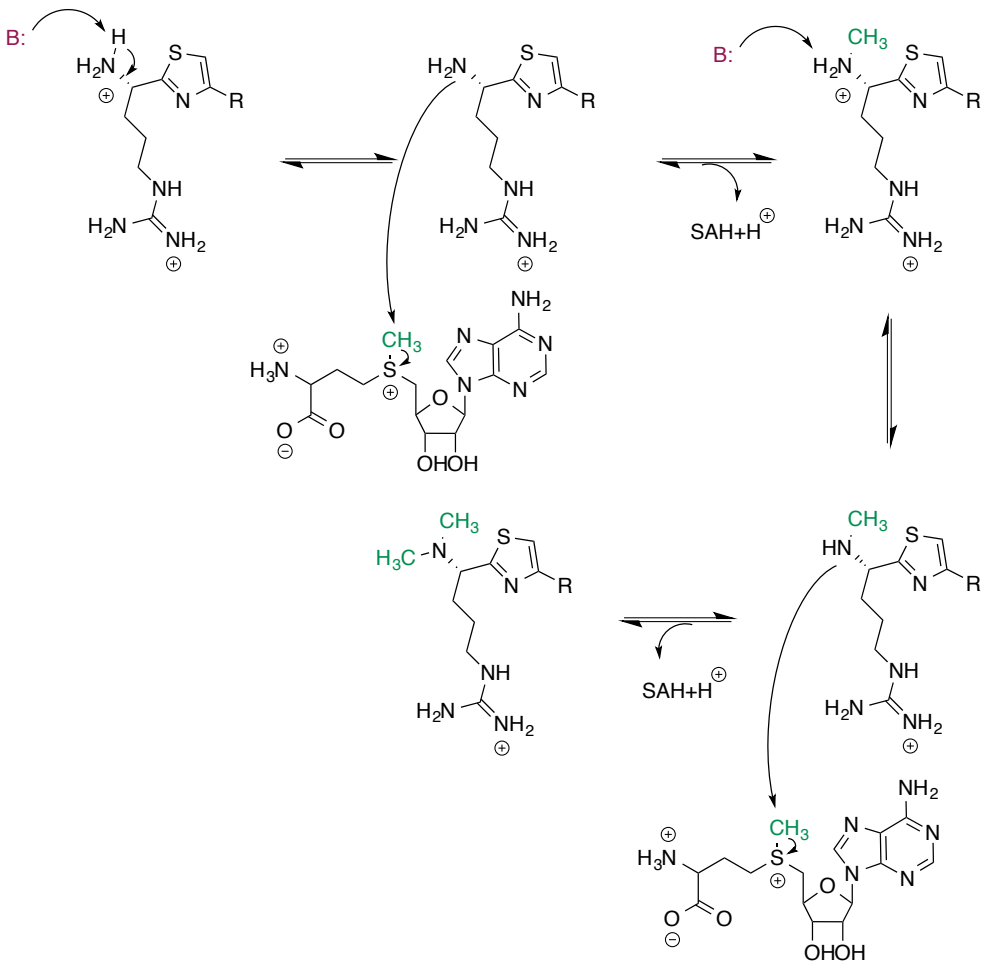

Fig. 3.10 Proposed mechanism of plantazolicin methyltransferase reaction. A base at active site deprotonates the α-amine of the N-terminal Arg of the polyazolic substrate and facilitates its nucleophilic attack onto the methyl group of SAM. Product SAH is then displaced by a second molecule of SAM. The monomethylated α-amine is further deprotonated and performs the second nucleophilic attack to form the dimethylated product.

Tables

Table 3.1 Steady-State Kinetic Parameters for BamL and BpumL with Truncated DesmethylPZN Analogs

BamL	K_M (10^{-6} M)	k_{cat} (10^{-2} s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
Arg-Az_1	26.8 ± 3.3	9.3 ± 0.4	3.46×10^3
Arg-Az_3	18.3 ± 0.5	10.5 ± 0.1	5.75×10^3
Arg-Az_5	14.6 ± 2.0	11.0 ± 0.4	7.53×10^3
Arg-Az_3 (BpumL)	22.5 ± 2.3	14.4 ± 0.7	6.38×10^3

Table 3.2 Steady-State Kinetic Parameters for BamL Variants against Arg-Az_3

BamL	k_{cat}/K_M (M^{-1} s^{-1})
Wild-type	57.5×10^2
T38F	3.3×10^2
R42A	4.7×10^2
L132F	3.5×10^2
L162F	1.2×10^2
Y182F	1.7×10^2

Chapter 4 Structural Studies on the Prenyltransferases in Cyanobactins Biosynthesis

Introduction

Natural products are good sources for drug discovery and provide chemists with scaffolds for further improvement or novel design. A majority of natural products are, however, challenging for synthesis and thus the exploitation and optimization of their original biosynthetic pathways are promising approaches for generating unnatural variants. Synthetic biologists have developed strategies for *in vivo* production of various classes of natural products and one substantially explored class is cyanobactins (Tianero, Donia, Young, Schultz, & Schmidt, 2012; Ruffner, Schmidt, & Heemstra, 2015). Cyanobactins are ribosomally synthesized and post-translationally modified peptides (RiPP) (Arnison, et al., 2013) and generally contain multiple heterocycles within a macrocyclic structure (Donia, et al., 2006; Donia, Ravel, & Schmidt, 2008; Sivonen, Leikoski, Fewer, & Jokela, 2010) (Fig. 4.1 A). The enzymatic mechanisms of heterocycle installation and macrocycle formation are thoroughly investigated and demonstrate strict dependency on the interaction between enzymes and recognition sequences harbored in the precursor peptide, similar to other typical RiPP modification reactions (McIntosh & Schmidt, 2010; McIntosh, Donia, & Schmidt, 2010; Agarwal, Pierce, McIntosh, Schmidt, & Nair, 2012; Koehnke J. , et al., 2012; Koehnke J. , et al., 2013; Sardar, Pierce, McIntosh, & Schmidt, 2015). In comparison, these tailoring enzymes are rather promiscuous towards the core peptides under modification, since cyanobactin precursor peptide naturally bears multiple copies or variants of core peptides (Donia, Ravel, & Schmidt, 2008; Donia & Schmidt, 2011). This substrate tolerance provides the opportunity for the establishment of a library of cyanobactin variants. Some cyanobactins further undergo additional modifications after macrocyclization, independent of the

recognition sequences. One commonly occurred decoration is prenylation with isoprenoids, such as dimethylallyl group or geranyl group (McIntosh, Donia, Nair, & Schmidt, 2011). This modification increases the lipophilicity of peptide products and could potentially confer other functions or preferable features as drug candidates. However, the substrate recognition and reaction mechanism of this process is still elusive.

Prenylation modification occurs in diverse biological processes of both prokaryotic and eukaryotic organisms. It is a common and important modification on eukaryotic proteins involved in essential cellular pathways. It facilitates membrane anchorage of the modified protein, alter its function and localization, and initiate cellular signaling transduction (Hannoush & Sun, 2010; Zverina, Lamphear, Wright, & Fierke, 2012). Most protein prenylations in eukaryotes involve farnesyl (C_{15}) or geranylgeranyl (C_{20}) group and the mechanisms of corresponding prenyltransferases have been comprehensively studied (Liang, Ko, & Wang, 2002; Palsuledesai & Distefano, 2015). Prenylation of small molecules is an essential biosynthetic step for some secondary metabolites, and for other metabolites, a crucial modification to confer bioactivity, product diversity and lipophilicity (Tello, Kuzuyama, Heide, Noel, & Richard, 2008; Heide, 2009). The small molecules undergoing modification and the types of isoprenoids used are more diverse than that in protein prenylation reactions. The corresponding prenyltransferases attract the attention of researchers due to the substrate promiscuity and potential biotechnological applications. Prenyltransferases that modify peptide substrates are less common and insufficiently investigated, though. One notable example is the *Bacilli* geranyl/farnesyl transferase ComQ involved in the biosynthesis of quorum-sensing peptide ComX (Magnuson, Solomon, & Grossman, 1994). It geranylates/farnesylates the indole ring of Trp similar to some small molecule prenyltransferases but displays strong dependency on

the sequence of flanking peptides (Tsuji, et al., 2012). The other example is the cyanobactin prenyltransferases discussed here, which select both the substrate residue and the flanking residues. Investigation on enzymatic mechanism and substrate preference is of great value in further expanding the cyanobactin library and generating non-cyanobactin peptides with isoprenoid modifications.

Genome mining of cyanobactin biosynthetic clusters led to the discovery of putative prenyltransferases with predicted isoprenoid-modified products isolated (Donia & Schmidt, 2011; Leikoski, et al., 2013). Despite the lack of sequence similarity with representative prenyltransferases, these proteins are postulated to utilize short isoprenoids (C_5 or C_{10}) and catalyze forward (attachment made through C1 atom) or reverse (attachment through C3 atom) prenylation on the hydroxyl groups of Tyr or Ser/Thr (Fig. 4.1 B). LynF from *Lyngbya aestuarii* was the first *in vitro* investigated cyanobactin prenyltransferase. The reported catalytic mechanism and substrate tolerance validated its function as an *O*-prenyltransferase for a group of Tyr containing peptides beyond its cognate substrate, even including small molecules boc-L-Tyr and boc-D-Tyr (McIntosh, Donia, Nair, & Schmidt, 2011). It is obvious that there is no such recognition sequence requirement for prenylation reaction but certain unknown determinants are definitely indispensable. TruF from trunkamide biosynthetic pathway was proposed to reverse-prenylate Ser/Thr (Donia, Ravel, & Schmidt, 2008). PirF from *Microcystis aeruginosa* PCC7005 was speculated to add geranyl group to its substrates (Leikoski, Fewer, Jokela, Alakoski, Wahlsten, & Sivonen, 2012). However, no further mechanistic studies were so far reported to explain their substrate preference. The prenyltransferase gene even appears in the *pat* cluster that produces no prenylated cyanobactins (Schmidt, et al., 2005). The structure of the putative prenyltransferase PatF was reported previously but provided limited information on its catalytic

deficiency with tested substrates (Bent, Koehnke, Houssen, Smith, Jaspars, & Naismith, 2013). Recent bioinformatic study revealed a new group of short, linear cyanobactins that are prenylated on N-terminal α-amine and methylated on C-terminal carboxyl group by a methyltransferase-prenyltransferase fusion protein, which displays a different substrate spectrum compared to the classic cyanobactin prenyltransferases (Leikoski, et al., 2013). The substrate selectivity and catalytic promiscuity of cyanobactin prenyltransferases allow these enzymes to be potential tools for enzymatic attachment of isoprenoids onto peptides. Therefore the mechanism and substrate tolerance study would be required.

To gain insights in the reaction mechanism and substrate specificity of cyanobactin prenyltransferases, co-crystal structures with prenyl donor analog and linear or macrocyclic substrates were determined for PagF in the prenylagaramide (*pag*) gene cluster from *Oscillatoria agardhii*. PagF reveals an α/β barrel structure reminiscent of the bacterial ABBA family aromatic prenyltransferases. Both of the prenyl donor and acceptor bind inside the barrel for catalysis. A peptide as prenyl acceptor is required to exclude solvent molecules from active site channel, prevent quenching the reactive carbocation intermediate, and ensure productive prenylation reaction. Primary sequence analysis on the annotated cyanobactin prenyltransferases explains the lack of activity for PatF in the patellamide gene cluster and provides clues on distinguishing functional prenyltransferases from incompetent ones. The putatively active ThcF in cyanothecamide gene cluster from *Cyanothece* sp. PCC7425 and PirF in piricyclamide gene cluster from *Microcystis aeruginosa* PCC7005 were tested for activity and confirmed as Tyr prenyltransferase and Tyr geranyltransferase, respectively. The native structure of PirF was determined to further justify that cyanobactin prenyltransferases adopt a solvent-exposed open channel to accommodate peptide substrates.

Experimental Procedures

Cloning, Protein Expression and Purification

The *pagF* gene from prenylagaramide (*pag*) gene cluster of *Oscillatoria agardhii* was cloned and ligated into pET-28b vector (Novagen). The resulting recombinant PagF contained a cleavable N-terminal hexahistidine tag (His-tag). This plasmid pET28-PagF was transformed into Rosetta 2 strain for expression. The culture first grew at 37 °C in LB medium supplemented with 50 μg/mL kanamycin and 25 μg/mL chloramphenicol until $O.D._{600}$ reached 0.5. Overexpression of His-PagF was induced with 0.3 mM IPTG at 18 °C for 20 h. Cells were pelleted by centrifugation at 3566 ×*g* for 25 min (SORVALL RC-3B Plus), resuspended with harvest buffer (20 mM Tris pH 8.0, 500 mM NaCl, 10% glycerol) and lysed with French Press (AVESTIN EmulsiFlex-C5). The supernatant was collected by centrifugation at 23645 ×*g* for 1 h (Beckman J2-21M/E) and loaded onto a 5 mL HisTrap FF column pre-equilibrated with wash buffer (20 mM Tris, pH 8.0, 1 M NaCl, 5% glycerol, 30 mM imidazole). After washing with the same buffer for 5 column volumes to remove impurities, the His-PagF was eluted with elution buffer (20 mM Tris, pH 8.0, 1 M NaCl, 5% glycerol, 200 mM imidazole) in a 0-100% linear gradient over 40 mL. Protein purity was checked with SDS-PAGE and clean fractions were pooled. Protein solution was then treated with thrombin at 4 °C overnight to remove the His-tag. The tag-free PagF was further purified by size exclusion chromatography with a HiLoad 16/600 Superdex 75 pg column (GE Healthcare Life Sciences). The protein was stable in gel filtration buffer (20 mM Tris pH 7.5, 100 mM KCl) and was concentrated for further use.

The *thcF* gene from cyanothecamide gene cluster of *Cyanothece* sp. PCC7425 was cloned and ligated into pET-28b vector (Novagen). The resulting recombinant ThcF contained a

cleavable N-terminal hexahistidine tag (His-tag). The plasmid transformation, heterologous expression and cell disruption procedures were the same as pET28-PagF. The His-ThcF was purified with immobilized metal ion affinity chromatography (IMAC) as described in PagF purification. The protein purity was checked by standard SDS-PAGE and clean fractions were collected. The His-ThcF was further purified by size exclusion chromatography with a HiLoad 16/600 Superdex 75 pg column (GE Healthcare Life Sciences). The protein was stable in gel filtration buffer (20 mM Tris pH 8.0, 500 mM KCl) and was concentrated for further use.

The codon optimized *pirF* gBlock gene fragments from *Microcystis aeruginosa* PCC7005 was purchased from Integrated DNA Technologies Inc. It was amplified by PCR and cloned into pET-28b vector between *Nde*I and *Xho*I sites. The resulting recombinant PirF contained a cleavable N-terminal hexahistidine tag (His-tag). The heterologous expression and protein purification procedures were the same as pET28-PagF, except that the final gel filtration buffer is 20 mM HEPES pH 7.5, 500 mM KCl. The protein was concentrated and flash-frozen in liquid nitrogen for storage.

Crystallization and Structure Determination

A sitting drop vapor diffusion method was used for robotic crystallization condition screening. The initial hits were further optimized with hanging drop vapor diffusion method. A typical hanging drop was set up with 1 μL of protein mixed with 1 μL of precipitant solution and was equilibrated over a well containing the same solution at 9 °C. HEPES buffer was avoided in PagF crystallization since the buffer molecule binds in the substrate-binding channel. Ball-like crystals of PagF were able to grow at a protein concentration of 21 mg/mL and in 20 mM Tris pH 7.5, 200 mM NaCl, 24% PEG3350. The crystal form didn't change until additives (Hampton

Research) were introduced in the hanging drops. The thin needle-like PagF crystals were equilibrated into crystallization condition supplemented with 12% (v/v) PEG3350 for protection before flash-frozen in liquid nitrogen. To obtain enzyme-substrate complex, the native PagF crystals were soaked in crystallization solution supplemented with 4 mM DMSPP, 10 mM $MgCl_2$, and 10 mM boc-L-Tyr, or 20 mM Tyr-Tyr-Tyr, or 100 μM cyclic[INPYLYP], for 4 h, 22 h, 41 h, respectively. All three peptides were gifts from Prof. E. Schmidt lab (University of Utah, Utah).

PirF crystals suitable for diffraction were obtained with hanging drop vapor diffusion method. 1 μL of 8 mg/mL PirF was mixed with 1 μL of precipitant solution (100 mM Bis-tris propane pH 7.0, 100 mM $(NH_4)_2SO_4$, 1 M sodium citrate tribasic dihydrate) and equilibrated over a well containing the same solution at 9 °C. The plate-like crystals were equilibrated into crystallization condition supplemented with 25% (w/v) threitol or trehalose for protection before flash-frozen in liquid nitrogen.

All of the data sets were collected at Sector 21 ID LS-CAT (Advanced Photon Source, Argonne National Labs, Illinois) and integrated and scaled using HKL2000 (Otwinowski & Minor, 1997) or XDS (Kabsch, 2010). Initial crystallographic phases of PagF were determined by single-wavelength anomalous diffraction method (SAD). A twelve-fold redundant data set from crystals soaked with trimethyl lead acetate was collected at lead absoption edge to a resolution of 2.65 Å. Heavy atom sites were located using HySS and refined using Phaser in the Phenix suite (Adams, et al., 2011), yielding an figure of merit (FOM) of 0.462. This initial map allowed the automatic building of most of the main chains and over half of the side chains using ARP/wARP. This model was refined against higher resolution data using REFMAC5 (Murshudov, et al., 2011) and further manually built using COOT (Emsley & Cowtan, 2004).

The free R factor was used to monitor building bias during building and crystallographic refinement (Kleywegt & Brünger, 1996). The diffraction data from crystals of PagF-ligand complexes were processed in a similar way. The phasess were determined using molecular replacement method with the refined ligand-free PagF structure as model (McCoy, Grosse-Kunstleve, Adams, Winn, Storoni, & Read, 2007). The resultant solutions were further automatically built using ARP/wARP, followd by rounds of manual rebuilding using COOT and refinement using REFMAC5. The diffraction data from native PirF crystal was integrated and scaled with autoPROC (Vonrhein, et al., 2011). The initial phase was determined using molecular replacement method with refined ligand-free PagF structure as model (McCoy, Grosse-Kunstleve, Adams, Winn, Storoni, & Read, 2007). The resultant solution was automatically built using Autobuild and further refined with Phenix.refine in the Phenix suite (Adams, et al., 2011), followed by manual building using COOT and refinement using REFMAC5.

Activity Analysis of ThcF and PirF

Overnight reactions were set up at room temperature with ThcF or PirF and appropriate substrates and supplements at following concentrations: 50 mM Tris pH 8.0, 200 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, 2 mM DMAPP/GPP, 0.2 μM enzyme, and 0.5 mM peptide substrate (YYY, FFPP, FFPC, AcYYY, AcYLYDVLPGTK, AY, VY, or LY). The overnight reaction was quenched with 50% methanol. 10 μL of each reaction was subject to positive-ESI-LC/MS using Agilent 1200 Series HPLC with a Alltech Denali C18 column (5 μm, 120 Å, 4.6 mm × 250 mm) coupled to Agilent LC/MSD Ion Trap XCT Plus mass spectrometer. Mobile phases consisted of water with 0.1% formic acid (solvent A) and methanol with 0.1% formic acid (solvent B). The

expected product and residual peptide substrate were analyzed using a linear gradient of 5-100% B over 20 min at a flow rate of 0.75 mL/min and monitored with their corresponding masses under single ion mode.

Results

Co-crystal Structures of PagF with Substrates

The broad substrate spectrum of cyanobactin prenyltransferases suggests they could prenylate any linear or cyclic peptide. In order to assess the structural determinants of substrate tolerance, we determined the crystal structures of native PagF (1.8 Å) and PagF in complex with prenyl donor analog dimethylallyl *S*-thiolodiphosphate (DMSPP) and boc-L-Tyr (2.1 Å), or Tyr-Tyr-Tyr tripeptide (1.9 Å), or a cognate substrate cyclic[INP**Y**LYP] (2.2 Å). The overall structure of PagF resembles the barrel architecture of the ABBA family of prenyltransferases with aromatic or indole-containing small molecules as substrates. It contains ten antiparallel β strands forming the inner face of the barrel and ten α helices surrounding outside (Fig. 4.2 A). Structural homology search of PagF against Protein Data Bank reveals prenyltransferases from the same family: fungal indole prenyltransferase FgaPT2 (PDB 3I4Z, RMSD 3.5 Å over 250 aligned Cα) from *Aspergillus fumigatus* (Metzger, et al., 2009), bacterial aromatic prenyltransferase NphB (PDB 1ZCW, RMSD 3.3 Å over 236 aligned Cα) (Kuzuyama, Noel, & Richard, 2005) and CloQ (PDB 2XLQ, RMSD 3.5 Å over 216 aligned Cα) (Metzger, Keller, Stevenson, Heide, & Lawson, 2010) from *Streptomyces* species. The inner β barrel of PagF aligns decently with those of FgaPT2, NphB and CloQ, but the orientation of α helices is different. A second distinct strucutural feature of PagF is that the β barrel forms a solvent exposed channel with open bottom, yet the typical ABBA family small molecule

prenyltransferases block the bottom of the channel with either an α helix (NphB and CloQ) or loops with aromatic residues (FgaPT2) (Fig. 4.3 A-D). It suggests the channel of PagF stays in an open state in the absence of substrates, but to ensure productive catalysis and avoid undesired hydrolysis of prenyl donor, additional bulky structures have to be present to exclude solvent and close the bottom.

Similar to typical ABBA family prenyltransferases, the active site of PagF is located in the channel. The co-crystal structures with different peptide substrates demonstrate that the binding of Tyr containing substrate blocks the bottom of the channel by pluging in the prenyl acceptor Tyr (Fig. 4.2 C, D). A hydrophobic environment is consequently created at the center of the channel for the dimethylallyl group of the prenyl donor. The diphosphate group of DMSPP is coordinated by R65, K136, H138, Y186, R288 and Mg^{2+} (Fig. 4.2 B). An active site divalent cation is typically required for prenylation reaction through coordinating the diphosphate oxygens and facilitating the formation of the electrophilic carbocation. Likewise, the function of cyanobactin prenyltransferases is dependent on the presence of Mg^{2+}. Canonical prenyltransferases generally contain a characteristic (N/D)DxxD sequence motif for coordinating active site Mg^{2+}, whereas Mg^{2+} binds in the active site of PagF without the signature motif. Lastly, the side chain of F222 also moves to avoid steric clash with DMSPP upon its binding.

Substrate Binding Site of PagF

A group of hydrophobic residues line the inner side of the PagF β barrel and create a hydrophobic environment to exclude solvent molecules from the active site channel. The placement of the substrate Tyr shields the environment and avoids unproductive quenching of the reactive intermediates. However, free amino acids are unable to completely close the channel.

A fairly large substrate is required to plug the bottom of PagF in a similar manner to the loops and helix at the bottom of the small molecule prenyltransferases. This provides the rationale for the observation that cyanobactin prenyltransferases only function on peptides or amino acids with bulky N-terminal modifications (Fig. 4.4 A-C). The carbonyl group of Tyr is in hydrogen-bonding distance with R140. The residue immediately C-terminal to Tyr is secured by van der Waals interactions with Q190 and L220, which suggests aliphatic/aromatic residue is preferred at this position for an accepted substrate (Fig. 4.4 D). A hydrophobic wall consisting of Y269, W271 and Y292 interacts with the residues N-terminal to Tyr and creates a kink in the peptide substrate (Fig. 4.4 D). Interestingly, the Tyr favored at the active site in the co-crystal structure is the very N-terminal one for Tyr-Tyr-Tyr substrate but neither of the other two (Fig. 4.4 B). Similarly, the favored prenyl acceptor Tyr in the co-crystal structure with cyclic[INP**Y**LYP] substrate is also the relatively N-terminal one immediately following Pro (Fig. 4.4 C). One possible explanation could be the aforementioned hydrophobic wall introduces steric hindrance for the positioning of the residues N-terminal to acceptor Tyr and thus limits the conformations that these residues have to adopt. But these conformations are not favored energetically. Implied from the preference of PagF over the residues N-terminal and C-terminal to Tyr, a linear peptide with the scaffold of a N-terminal Tyr immediately followed by an aliphatic/aromatic residue will be an effective substrate of PagF.

Notably, the superimposition of the three co-crystal structures with the cyclic substrate or either of the two linear substrates reveals almost identical positioning of the prenyl acceptor Tyr, sandwiched between W271 on one side and L67 and F69 on the other side (Fig. 4.4 E). The phenolic oxygen of Tyr is coordinated by E51 at a distance of 2.6 Å and is likely deprotonated for the subsquent nucleophilic attack onto the dimethylallyl group. The C1 atom of dimethylallyl

group is located directly above the phenolic oxygen of Tyr at a distance of 3.1 Å and is therefore the favored atom for the reaction rather than the C3 atom, which otherwise would result in a reverse prenylation. Similar distances and substrate orientations have also been observed between the dimethylallyl C1 atom and reactive substrate atom in NphB and FgaPT2 active sites, with a distance of 4 Å and 3.1 Å, respectively (Kuzuyama, Noel, & Richard, 2005; Metzger, et al., 2009). The relative positioning of substrate Tyr and the dimethylally group indicates that the planar allylic carbocation resulted from the cleavage of DMAPP will be stabilized through cation-π interaction with the Tyr aromatic ring.

Sequence Analysis of Cyanobactin Prenyltransferases

Biochemical characterization of cyanobactin prenyltransferases demonstrated several members incapable of catalyzing prenylation reaction either *in vivo* or *in vitro*, such as PatF in *pat* pathway and TenF in *ten* pathway. *tru* pathway encodes two putative prenyltranferases TruF1 and TruF2, but only the former one showed reverse prenylation activity. The reason for the catalytic deficiency of these prenyltransferases is still elusive but sequence comparison with active prenyltransferases might provide some clues. Primary sequence alignment of the cyanobactin prenyltransferases indicates that the key residues coordinating the diphosphate moiety of the prenyl donor in PagF co-crystal structures may play a crucial role in enzyme activity. In the characterized active prenyltransferases, the R65, K136, H138, Y186 and R288 (in PagF numbering) are highly conserved; in contrast, more than one of these residues are absent in the known inactive prenyltransferases (Fig. 4.5 A). In PatF, K136 is Met, H138 is Tyr, Y186 is Gly, and R288 is Cys. In TenF, R65 is Asp, H138 is Tyr, Y186 is Leu, and R288 is Ser. In TruF2, R65 is Gln, H138 is Ser, Y186 is Gly, and R288 is Leu. It seems that this diphosphate-

binding motif signifies the catalytic capability or deficiency of a cyanobactin prenyltransferase. However, examination of either the primary sequences or the limited structural information on cyanobactin prenyltransferases didn't provide much insight on the preference for forward or reverse prenylation or on the substrate selectivity.

To gain more insights in sequence-function relationship, a Sequence Similarity Network of annotated standalone cyanobactin prenyltransferases was built based on all-by-all protein sequence BLAST results. The network was analyzed at the cutoff E-value of 1e-80 and proteins with higher sequence similarity cluster together (Fig. 4.5 B). All the known inactive prenyltransferases, PatF, TenF, and TruF2, cluster together, suggesting others in the same cluster might not be active as well. LynF and TruF1 appear as singletons in the network, which is consistent with their unique functions among all the characterized cyanobactin prenyltransferases. LynF is the only known prenyltransferase to catalyze reverse *O*-prenylation on Tyr to date; similarly, TruF1 is the only known one to catalyze the same reaction on Ser/Thr. The cluster where PagF is located potentially represents the ones catalyzing forward *O*-prenylation on Tyr, but none except PagF has been investigated *in vitro*. PirF is encoded in the *pir* gene cluster from *Microcystis aeruginosa* PCC7005 and production of geranylated instead of prenylated cyanobactins has been detected (Fig. 4.1 A). It suggests PirF is an active geranyltransferase *in vivo* and the Tyr or Ser residue in the cyclic peptide is the potential site for geranyl group attachment (*pir* cluster lacks the heterocyclase gene for azoline ring installation and thus Ser is not modified) (Leikoski, Fewer, Jokela, Alakoski, Wahlsten, & Sivonen, 2012). Nevertheless, there are also examples with no reported *in vivo* prenyltransferase activity in this "active protein" cluster. For example, ThcF is encoded in cyanothecamide (*thc*) gene cluster but only unmodified cyclic cyanobactins were isolated despite the presence of Tyr residue available

for prenylation (Donia & Schmidt, 2011; Houssen, et al., 2012). Therefore ThcF is possibly inactive *in vivo*. As a consequence, a question mark is put on the function assignment of the protein cluster represented by PagF.

Biochemical Characterization of ThcF and PirF Activity

In order to assess the proposed Tyr prenylation activity of the protein cluster represented by PagF, the recombinantly expressed "inactive" ThcF and the "active" PirF were tested for prenylation/geranylation activity with a variety of peptide substrates and the modified products were monitored by LC/MS. In the presence of excess amount of Mg^{2+} and prenyl group donor DMAPP, ThcF is able to singly prenylate Tyr-Tyr-Tyr (YYY) tripeptide but doesn't effectively attach the second prenyl group (Fig. 4.6). To verify that the only prenyl acceptor is the Tyr side chain hydroxyl group instead of the N-terminal α-amine, the Phe-Phe-Pro-Pro (FFPP) peptide was used for enzymatic assay to mimic a Tyr containing peptide with the reactive N-terminal α-amine but not the side chain hydroxyl group. As expected, ThcF showed no activity towards α-amine, indicating its substrate specificity towards Tyr containing peptide. Similarly, a Tyr-Tyr-Tyr tripeptide with acetylated α-amine (AcYYY) was also successfully modified by ThcF, but at a considerably lower yield compared with the YYY tripeptide containing free α-amine, which indicates that possible steric hindrance introduced by the *N*-acetylation modification affects the accommodation of substrate peptide in the enzyme active site. The *in vitro* or *in vivo* activity of ThcF towards its native substrates wasn't evaluated here to explain the detection of only unprenylated cyanobactins from the native producer. There is the possibility that the active site is not able to accommodate the size or the conformation of the native cyclic peptide substrate and this could be described as incomplete co-evolution of enzyme and its cognate substrates. A

second possibility is the prenylated cyanobactins were actually produced but only in small amount beyond the detection limit, or were lost during isolation process.

The geranyltransferase activity of PirF was assayed with varied peptide substrates in a similar manner to ThcF and the modified products were monitored by LC/MS (Fig. 4.7). With excess amount of Mg^{2+} and geranyl diphosphate (GPP), PirF is able to effectively geranylate the YYY tripeptide. With GPP substituted by the C_5 donor DMAPP in the activity assay, no modification was observed on the same tripeptide, which implies its strong selectivity towards isoprenoid group donor. To verify PirF catalyzes Tyr prenylation rather than *N*-prenylation on α-amine, a Phe-Phe-Pro-Cys (FFPC) peptide substrate was tested with PirF and as predicted didn't trigger geranylation reaction. A linear peptide Tyr-Leu-Tyr-Asp-Val-Leu-Pro-Gly-Thr-Lys with *N*-acetylation on α-amine (AcYLYDVLPGTK) was also successfully geranylated but with low conversion ratio. Similar to the observation in ThcF reaction, the extra structure N-terminal to the acceptor Tyr might introduce steric hindrance for proper substrate orientation in the active site. To access the effect of N-terminal structure on catalysis, PirF activity was evaluated against a panel of Tyr containing dipeptide with different N-terminal aliphatic residues. With Ala or Val N-terminal to the acceptor Tyr, PirF is able to completely convert the peptides in the assay condition. With Leu at N-terminus, the conversion ratio is significantly compromised, indicating a relatively inefficient catalysis by PirF. These results demonstrate that PirF catalyzes geranylation reaction on Tyr residue and the reaction is negatively affected by the bulky structure N-terminal to the acceptor Tyr.

Structural Characterization of PirF

To gain structural insights on PirF, the crystal structure of native PirF is determined at a resolution of 2.46 Å. The overall structure of PirF resembles the barrel structure of the previously described PagF structure, composed of ten antiparallel β strands as the core of the barrel surrounded by ten α helices (Fig. 4.8 A). Residues spanning M1–S10 and F223–N229 were not modeled in the structure due to the lack of electron density. The superimposition of PirF structure with PagF reveal the same topology and orientation of secondary elements (Fig. 4.8 B). Similar to PagF, the central barrel of PirF also forms an open solvent-accessible channel with no auxiliary secondary structure blocking either end, suggesting the utilization of peptide substrate to plug the channel. The nearly perfect superimposition of PirF and PagF structures implies that the cyanobactin prenyltransferases are likely to use the same open-channel strategy to accommodate the bulky peptide substrates. Upon substrate binding, a closed hydrophobic environment is consequently created, free of solvent molecules and suitable for productive catalysis.

Superimposition of PirF with PagF reveals that the residues coordinating the DMSPP and the acceptor Tyr are conserved in the active site channel of PirF. PirF residues R65, K134, H136, Y184 and R286 are located in the identical positions compared to their counterparts in PagF (Fig. 4.9 A), suggesting their role in binding the diphosphate moiety of geranyl group donor GPP. The residues involved in binding the acceptor Tyr in PagF structure are also observed in similar positions in PirF structure (Fig. 4.9 B). M67 and F69 forms one side and W269 forms the opposite side. The acceptor Tyr is likely sandwiched between them similar to the Tyr coordination in PagF. PirF residue E51 is superimposable with the proposed active site base E51 in PagF and is thus expected to deprotonate the acceptor Tyr to facilitate its nucleophilic attack onto the geranyl carbocation cleaved from the diphosphate. The similar residue arrangement of

PirF and PagF in the acceptor Tyr-binding site implies that this might be a characteristic organization of the active site for Tyr substrate. Scrutinization of the primary sequence showed that these residues are relatively conserved in the known Tyr cyanobactin prenyltransferases. In contrast, in TruF1, the only known Ser/Thr prenyltransferase to date, short polar residues appear at these positions.

A distinct difference between PirF and PagF is that PirF strictly uses C_{10} isoprenoid donor (GPP) but not C_5 donor (DMAPP). Therefore, a bigger pocket for the isoprenoid donor accommodation is expected in PirF compared to PagF. A close look at the hydrophobic residues in the vicinity of diphosphate coordination site led to a possible explanation for the preference of C_{10} isoprenoid donor (GPP). Lack of the bulky F222 (PagF numbering) in PirF in conjunction with the outward twist of the β7 strand creates extra space in the active site channel (Fig. 4.9 C), which might favor the placement of geranyl group. This change is not resulted from missing the Phe in the primary sequence but caused by a Gly insertion before the Phe (Fig. 4.9 D). The Gly forces the succeeding Phe to face the outside of the barrel rather than introduce more steric hindrance inside the active site channel. However, this observation still could not perfectly answer why the C_5 donor (DMAPP) is not accepted in a more spacious environment in PirF. It is possible there are other interactions between the geranyl group and the inner face of the barrel, which the smaller C_5 could not fulfill. It is also possible that the C_5 donor (DMAPP) does bind in the active site channel but it fails to prevent the solvent from quenching the reactive carbocation due to its smaller volume. In fact, this Gly insertion also appears in ThcF (Fig. 4.5 A), which is apparently active towards DMAPP (Fig. 4.6). Although ThcF wasn't tested with GPP, the Gly insertion and the spacious environment are probably not the only determinants for the preference of isoprenoid donors.

Discussion

Cyanobactins adopt several strategies to increase their structure rigidity, stability and hydrophobicity, such as the heterocyclization of Ser/Thr/Cys, macrocyclization from a linear peptide and prenylation on the residual polar residues Ser/Thr/Tyr. While the heterocyclization and macrocyclization mainly increase the rigidity and resistance to nonspecific hydrolysis, prenylation with dimethylallyl or geranyl group considerably improves the lipophilicity. The involved enzymes, cyanobactin prenyltransferases, are among the few examples in bacteria that use peptides as substrate. Despite their independence of leader peptide, they do show high specificity towards the isoprenoid donor and the acceptor residue, making them potential tools for residue-specific prenylation with variable flanking sequences.

To investigate the determinants of substrate specificity, co-crystal structures of prenyltransferase PagF with linear substrates or the native cyclic substrate were determined. The α/β barrel structure of PagF resembles the ABBA family aromatic prenyltransferases. But unlike the characterized small molecule ABBA prenyltransferases, the PagF barrel is fully solvent exposed with no blockage at either end. The structure of the piricyclamide geranyltransferase PirF presented here showed almost identical overall structure. It suggests that cyanobactin prenyltransferases share the same open-channel structure, consistent with their function as peptide prenyltransferase. The bulky peptide substrate serves as the plug of the open channel and assists in creating a hydrophobic environment for productive catalysis. Both the prenyl donor and acceptor are located inside the active site channel in PagF co-crystal structures. The diphosphate moiety of prenyl donor is coordinated through a hydrogen-bonding network from the polar side chains as well as the Mg^{2+}, with dimethylallyl group positioned in a hydrophobic environment.

The peptide substrate is also harbored in a hydrophobic environment with the acceptor Tyr held in proximity to the dimethylallyl group. The Tyr stabilizes the carbocation intermediate through π-cation interaction with the dimethylallyl group cleaved from diphosphate. The phenolic oxygen of Tyr is located directly beneath the C1 atom of dimethylallyl group in PagF and the forward *O*-prenylation is thus favored rather than the reverse prenylation.

Sequence analysis of the annotated cyanobactin prenyltransferases supports the significance of the residues involved in substrate coordination and partially explains the substrate specificity. The diphosphate-binding residues identified in PagF co-crystal structures are also observed at identical positions in PirF and are highly conserved in the known active cyanobactin prenyltransferases. The substitution by other residues is found in all known inactive ones. The hydrophobic residues sandwiching the acceptor Tyr in PagF and potentially in PirF are only relatively conserved in Tyr prenyltransferases, consistent with their substrate specificity to Tyr containing peptides. These residues are replaced by short polar residues in TruF1, the only identified Ser/Thr cyanobactin prenyltransferase to date. These substitutions, probably together with other active site residues, establish a favorable local environment for Ser/Thr binding instead of the aromatic Tyr. The PagF residue E51 in hydrogen-bonding interaction with the Tyr phenolic oxygen appears in all active cyanobactin prenyltransferases and is likely the universal active site base for deprotonating hydroxyl group. The factors discriminating forward versus reverse prenylation remain elusive and unpredictable from primary sequence, but the relative orientation of the reactive oxygen and the isoprenoid donor might be an explanation. When the oxygen is located closer to the C1 atom of the dimethylallyl group, a forward prenylation occurs. Similarly, if the oxygen is in shorter distance to the C3 atom, a reverse prenylation is triggered.

The structures of PagF-substrate complexes and the activity assays for ThcF and PirF indicate that the Tyr cyanobactin prenyltransferases prefer peptide substrates with no bulky group immediately N-terminal to Tyr. Small protecting groups (such as boc and acetyl groups) and short aliphatic residues (such as Pro, Ala) are acceptable at the position N-terminal to Tyr. But substrates with bulkier substitutions are substantially blocked by the steric clashes in the substrate-binding tunnel. Consequently, the best peptide substrate seems to be one containing N-terminal Tyr. This N-Tyr scaffold satisfies the chemoselectivity and regioselectivity of the Tyr cyanobactin prenyltransferases described here and could serve as a general guideline for designing peptide substrates for these enzymes. Since the Tyr cyanobactin prenyltransferases are selective towards only one residue and quite promiscuous to the rest of peptide sequence, they are better tools for introducing isoprenoid modifications in peptides compared to the protein prenyltransferases that recognize a four-residue sequence.

With an increasing number of genomes sequenced from cyanobacteria, more cyanobactin prenyltransferases are discovered with varying substrate selectivity and are thus potential toolkit for introducing lipophilic isoprenoids to peptides. An example is the methyltransferase-prenyltransferase fusion protein (MT-PT) in the biosynthetic gene clusters of short linear cyanobactins. These clusters encode the macrocyclase but the peptide products are not cyclic, possibly due to a peptide-length requirement for the macrocyclase to function properly. To increase the hydrophobicity of the peptide products and protect the N- and C- termini from nonspecific proteolysis, the MT-PT prenylates the N-terminal α-amine and methylates the C-terminal carboxylate. This prenyltransferase domain is unique compared to the characterized standalone cyanobactin prenyltransferases, since it recognizes the α-amine instead of the residue side chains. Although the side chain of the N-terminal residue might still have to satisfy some

criteria, the specificity of this *N*-prenyltransferase to the backbone amine presents it as a promising tool for prenylating any peptide regardless of the sequence.

Figures

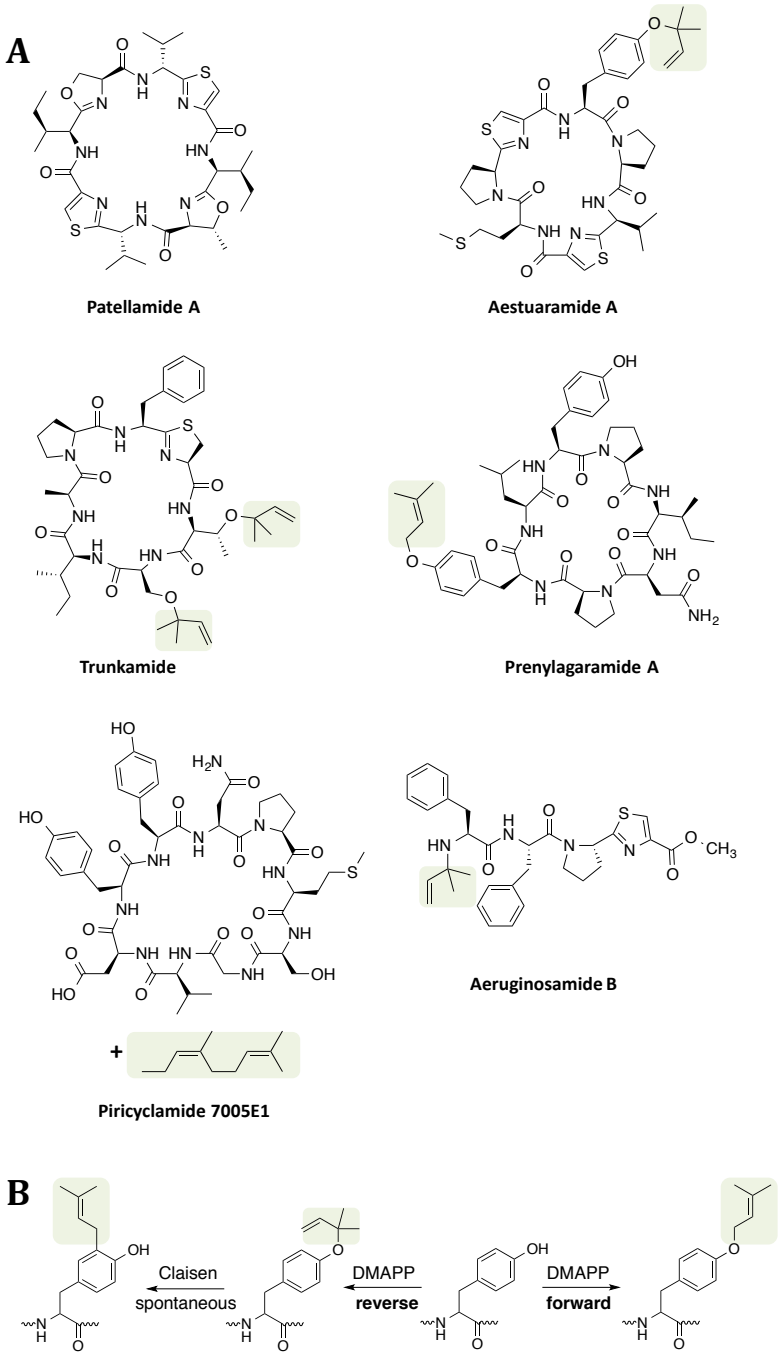

Fig. 4.1 **A.** Representative cyanobactins. Patellamide A is produced by the *pat* gene cluster of *Prochloron didemni* and contains no isoprenoid modification. Aestuaramide A is produced by

the *lyn* gene cluster of *Lyngbya aestuarii* CCY9616. The dimethylally group on Tyr hydroxyl group is the product of reverse prenylation and is further spontaneously converted to forward prenylation on the carbon atom at *meta* position. Trunkamide is produced by the *tru* cluster of *Prochloron didemni* and contains reversely prenylated Ser and Thr. Prenylagaramide A is produced by the *pag* gene cluster of *Oscillatoria agardhii* and contains a forward-prenylated Tyr. Piricyclamide 7005E1 is produced by the *pir* gene cluster of *Microcystis aeruginosa* PCC7005. It contains modification with geranyl group but the residue for attachment is not determined. Aeruginosamide B is a linear cyanobactin isolated from *Microcystis aeruginosa* PCC9432 with reverse prenylation on N-terminal α-amine and methylation at C-terminus. **B.** Scheme for forward and reverse prenylation on Tyr. The side chain oxygen of Tyr attacks the C1 atom of dimethylallyl group to result in forward *O*-prenylation. The nucleophilic attack on C3 atom of dimethylallyl group leads to reverse *O*-prenylation in the case of LynF reaction. Subsequent nonenzymatic Claisen rearrangement generates the forward *C*-prenylated product.

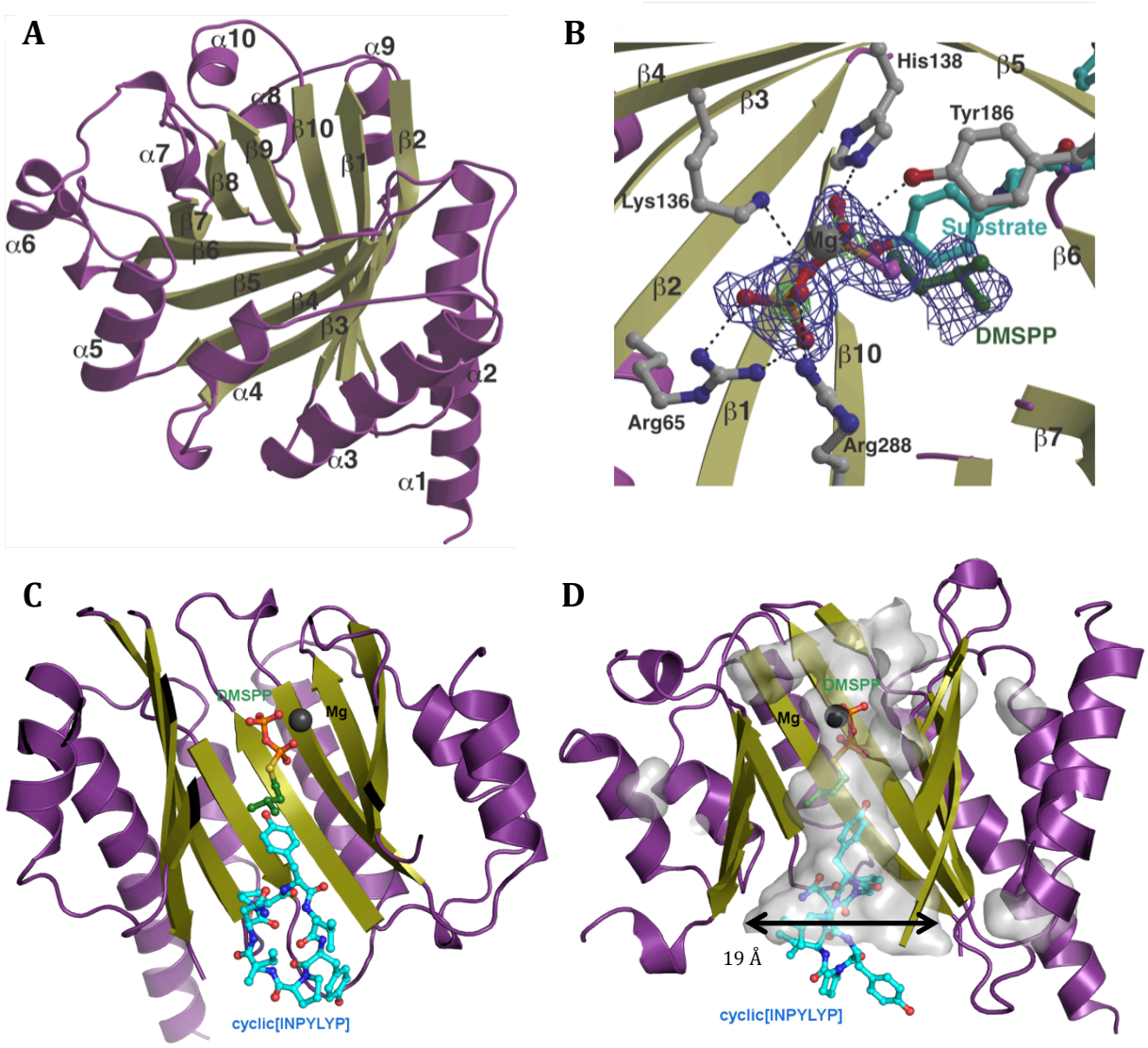

Fig. 4.2 **A.** Overall structure of native PagF. The β strands (colored in olive) comprising the central barrel and the flanking α helices (colored in purple) are numbered separately. **B.** Coordination of prenyl donor analog DMSPP (green sticks) in the substrate-binding channel. The diphosphate moiety of DMSPP is coordinated by PagF residues shown as gray sticks. Mg^{2+} ion is shown as gray sphere. Hydrogen bonds are shown as black dashed lines. The Tyr (cyan) from prenyl group acceptor is in proximity to the dimethylallyl group. The *Fo-Fc* omit map showing the electron density of DMSPP is contoured at 2.4σ (blue) and 10σ (green). **C.** Longitudinal

section of the PagF active site channel, showing the accommodation of both the prenyl donor analog DMSPP (green) and the native substrate cyclic[INP**Y**LYP] (cyan). The peptide substrate blocks the bottom of the channel. Mg^{2+} ion is shown as gray sphere. **D.** Longitudinal section of the PagF active site channel, showing the inner surface (gray) of the channel. The widest point of the channel is around 19 Å and is blocked by the bulky peptide substrate. Prenyl donor analog DMSPP is shown as green sticks. Native substrate cyclic[INP**Y**LYP] is shown as cyan sticks. Mg^{2+} ion is shown as gray sphere.

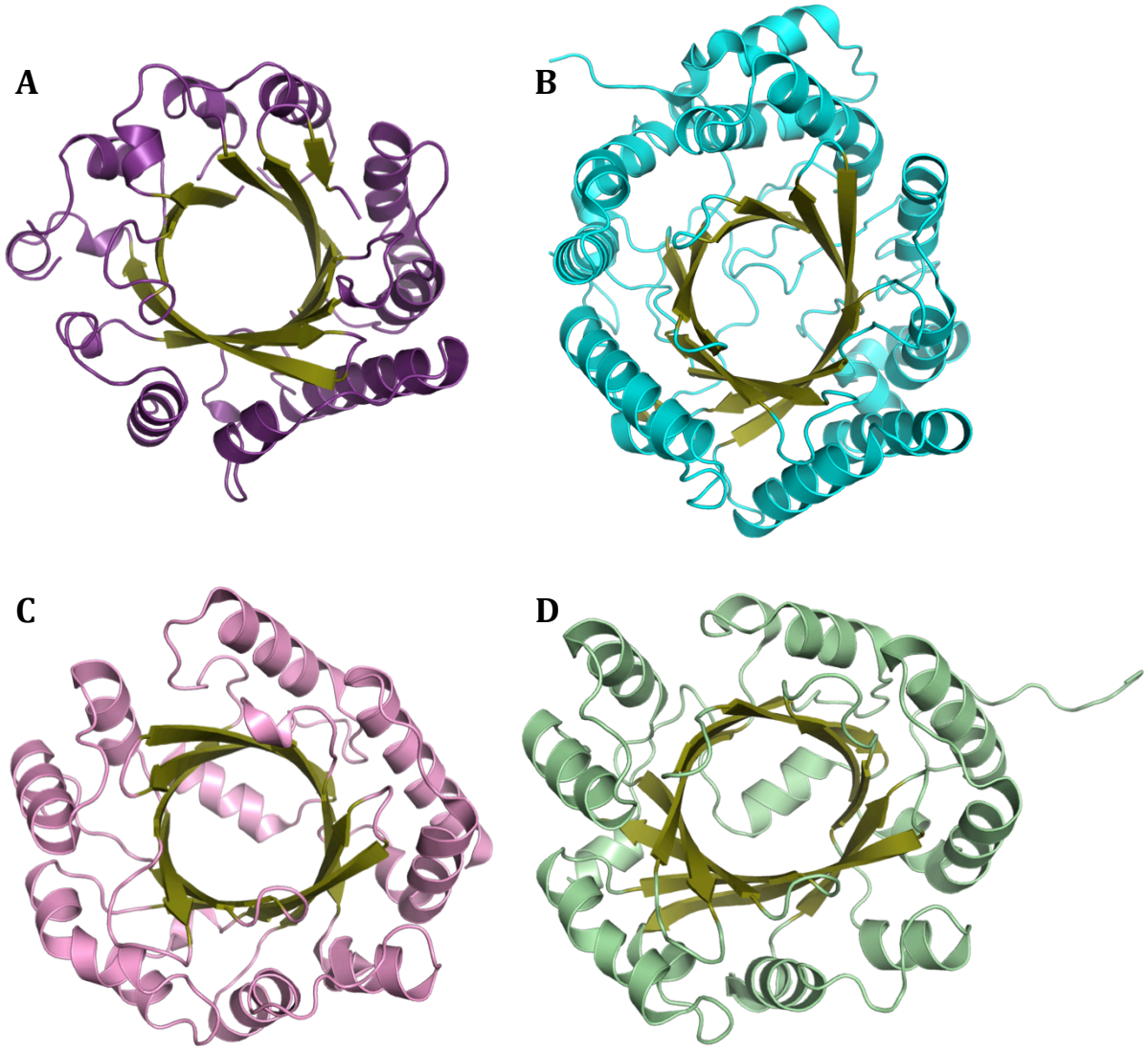

Fig. 4.3 **A.** A bird's-eye view of the PagF barrel structure. The barrel forms a solvent-accessible open channel. **B.** A bird's-eye view of FgaPT2 barrel structure. The bottom of the barrel is blocked by hydrophobic loops. **C.** A bird's-eye view of NphB barrel structure. The bottom of the barrel is blocked by an α helix. **D.** A bird's-eye view of CloQ barrel structure. The bottom of the barrel is blocked by an α helix.

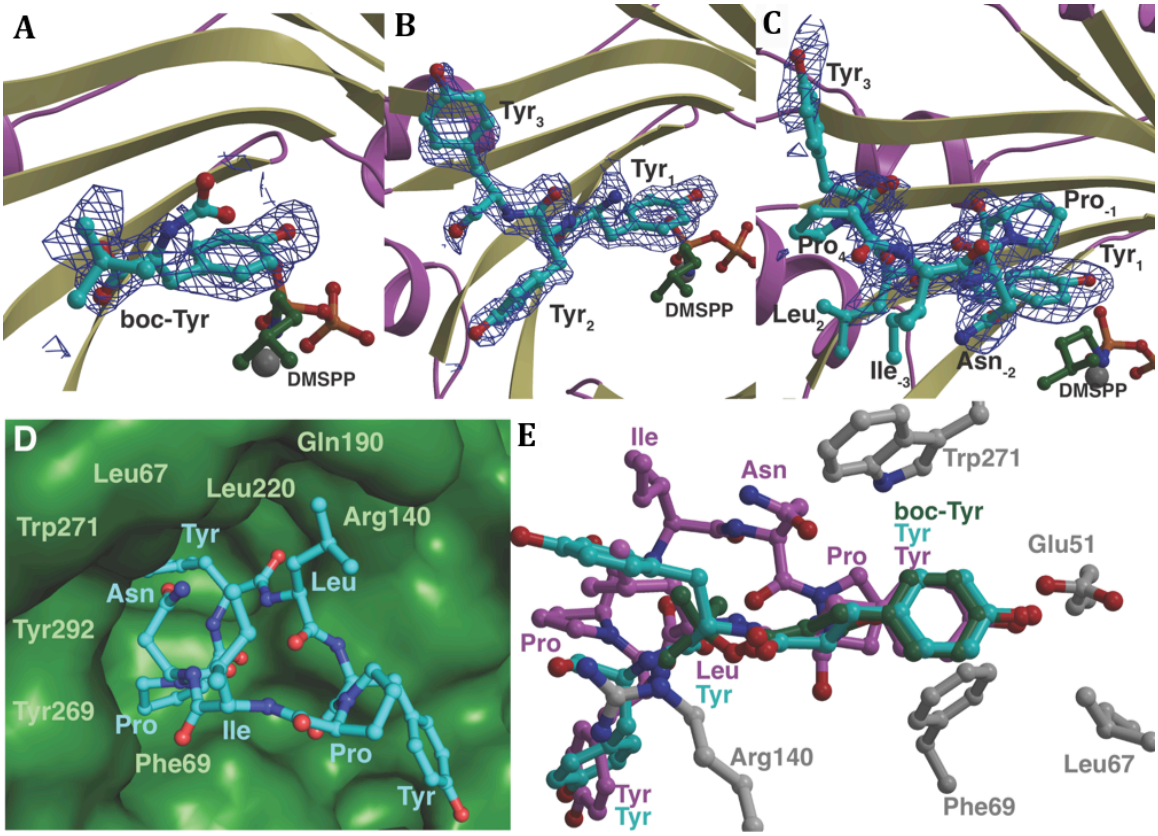

Fig. 4.4 **A.** A close view of the coordination of substrate boc-Tyr (cyan) in the substrate-binding channel. **B.** A close view of the coordination of substrate Tyr-Tyr-Tyr (cyan) in the substrate-binding channel. **C.** A close view of the coordination of substrate cyclic[INP**Y**LYP] (cyan) in the substrate-binding channel. **D.** A close view of the cyclic[INP**Y**LYP] (cyan) substrate situated in the hydrophobic chamber. **E.** Superimposition of the Tyr containing substrates boc-Tyr (green), Tyr-Tyr-Tyr (cyan) and cyclic[INP**Y**LYP] (purple) in the substrate-binding channel. The prenyl acceptor Tyr residues align identically between L67, F69 and W271.

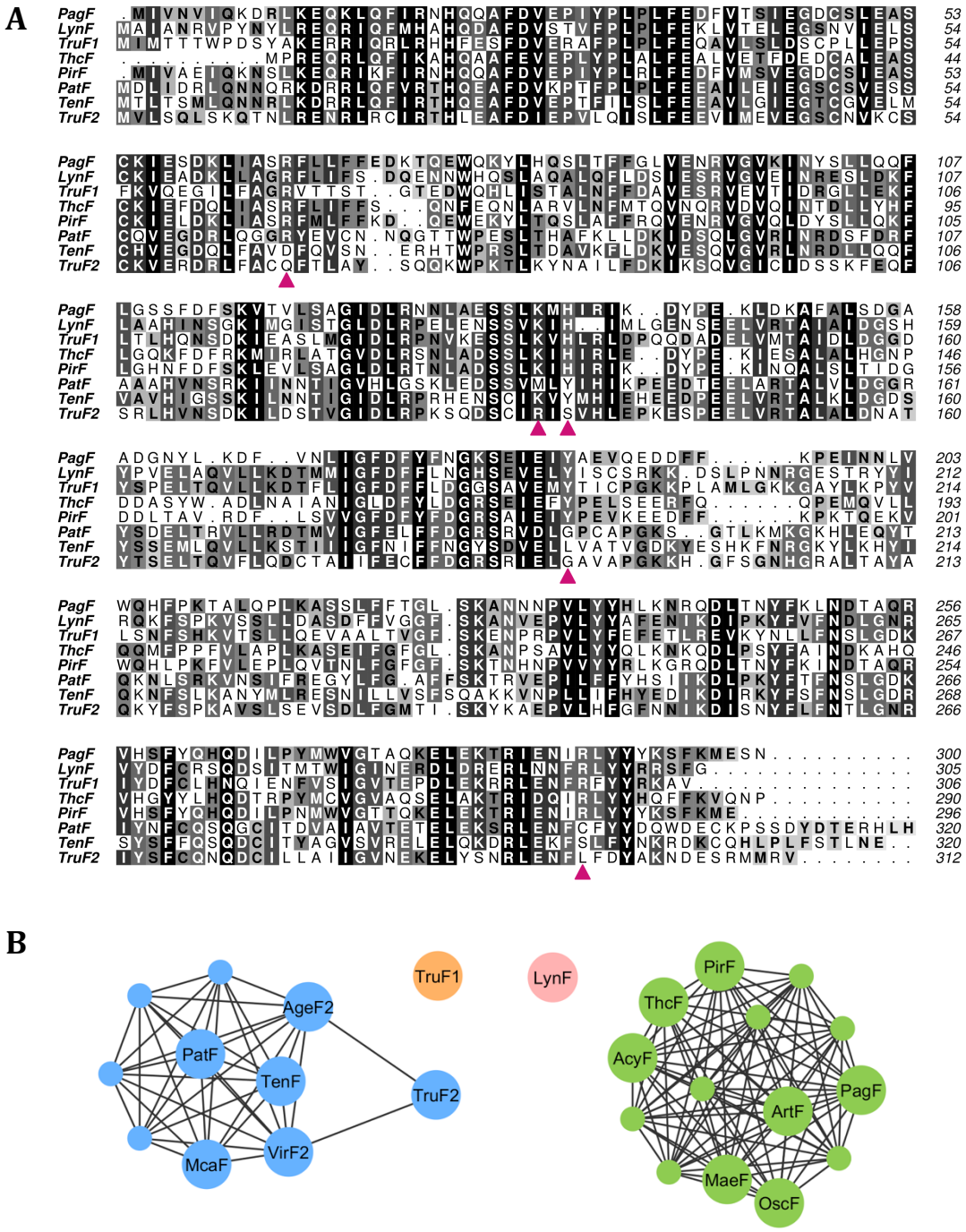

Fig. 4.5 **A.** Primary sequence alignment of representative annotated cyanobactin prenyltransferases. Residues coordinating diphosphate moiety of DMAPP in PagF structure are indicated with magenta arrowheads. **B.** Sequence Similarity Network (Gerlt, et al., 2015) of the

annotated standalone cyanobactin prenyltransferases displayed at 1e-80 in Cytoscape (Shannon, et al., 2003). The PagF sequence was input into the EFI-EST server as the query sequence to achieve the all-by-all BLAST result of sequences sharing a similarity not less than 1e-5. The result was displayed in Cytoscape with a sequence-length limit set to avoid fragmented sequences or sequences from multi-domain proteins. The E-value was set to 1e-80 (30-40% sequence similarity) for further analysis. Nodes in the same cluster show higher sequence similarity to each other compared to the similarity to nodes outside the cluster. Representative nodes are labeled with the name of the protein.

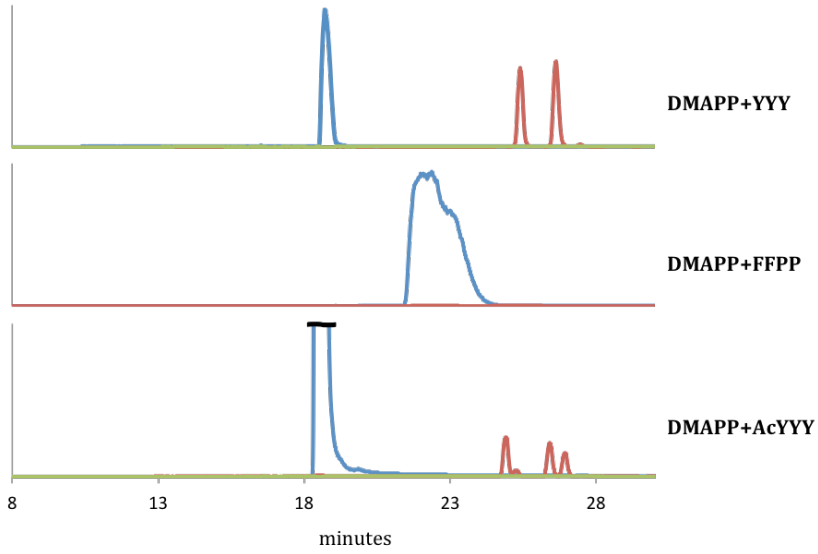

Fig. 4.6 LC/MS analysis of the ThcF prenylation reactions. The protential prenyl accepting peptide YYY, FFPP or AcYYY is incubated with ThcF and excess amount of DMAPP overnight. The expected mono- and di- prenylated products as well as the starting material are analyzed with LC/MS. The blue trace represents the starting material left in the reaction mixture. The red trace represents the expected mono-prenylated product. The green trace shows the di-methylated product.

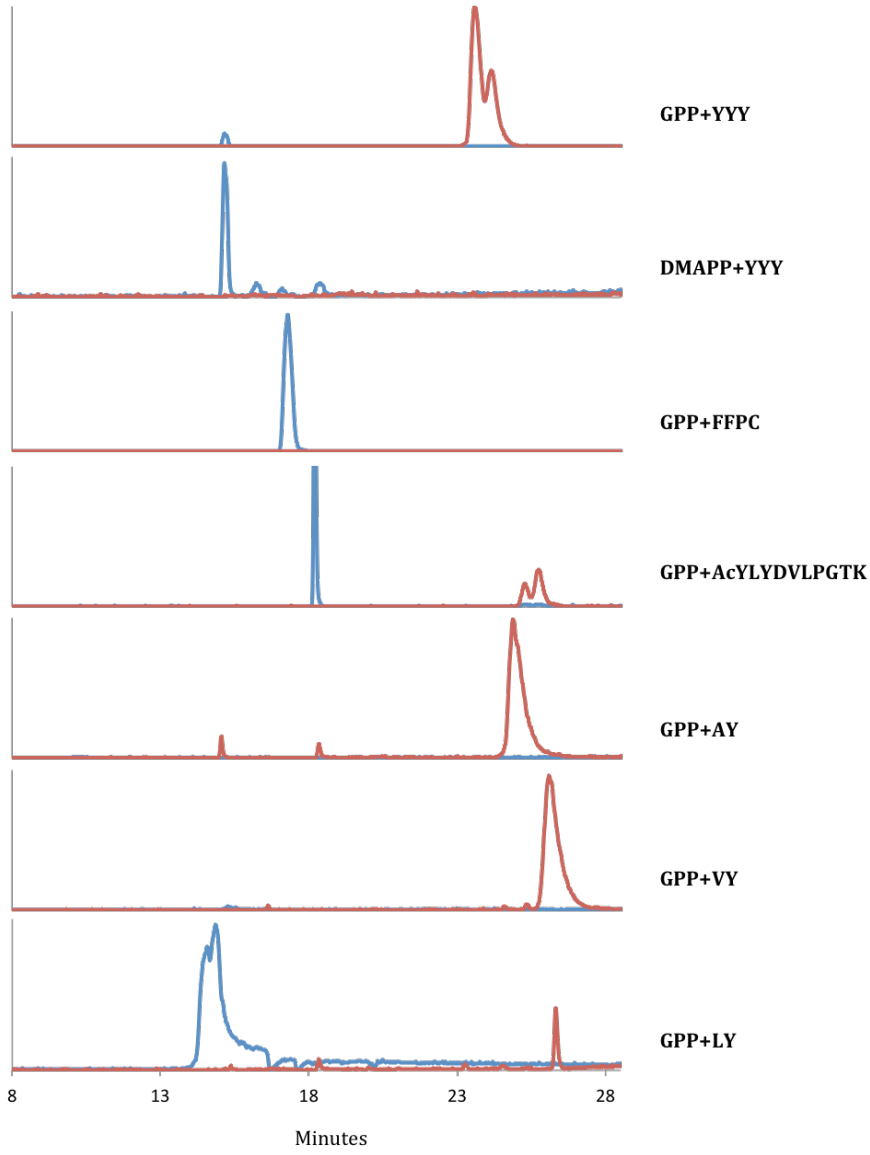

Fig. 4.7 LC/MS analysis of the PirF geranylation reactions. The protential geranyl accepting peptide YYY, FFPC, AcYLYDVLPGTK, AY, VY or LY is incubated with PirF and excess amount of GPP overnight. The expected mono-geranylated product (red trace) is analyzed with LC/MS as well as the starting material left (blue trace) in the reaction mixure. PirF is also tested for prenylation activity with DMAPP and peptide YYY.

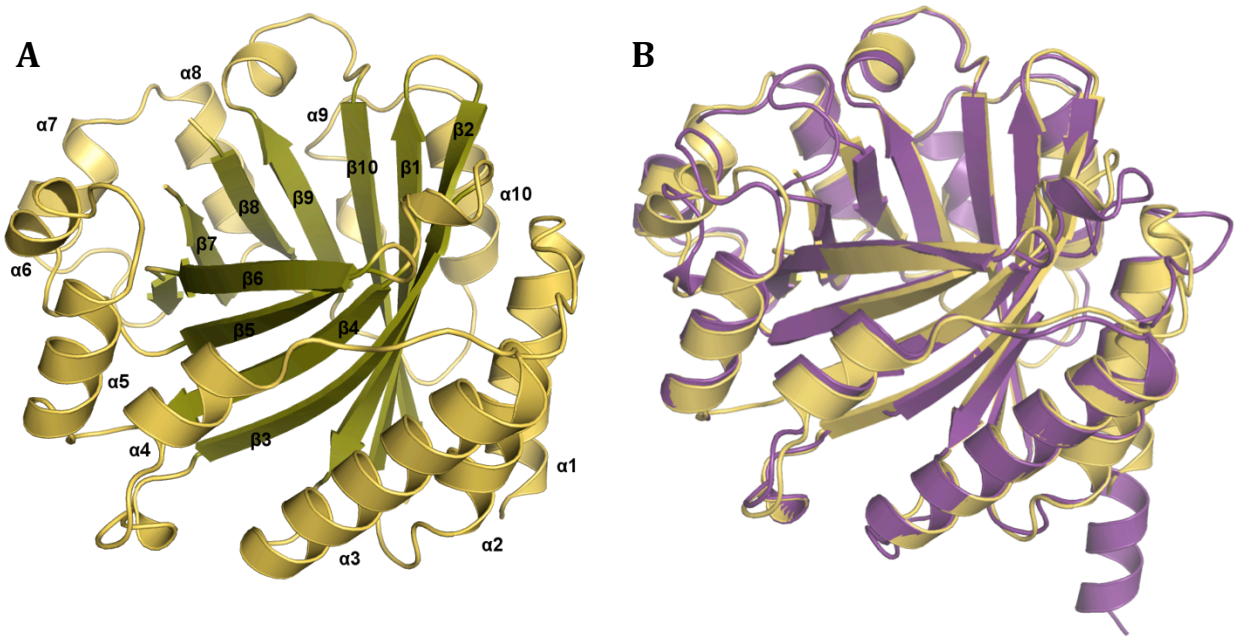

Fig. 4.8 **A.** Overall structure of native PirF. The β strands (colored in olive) comprising the central barrel and the flanking α helices (colored in gold) are numbered separately. **B.** Structural superimposition of PagF (purple) and PirF (gold). The central barrel of PirF forms a solvent-accessible open channel as in PagF.

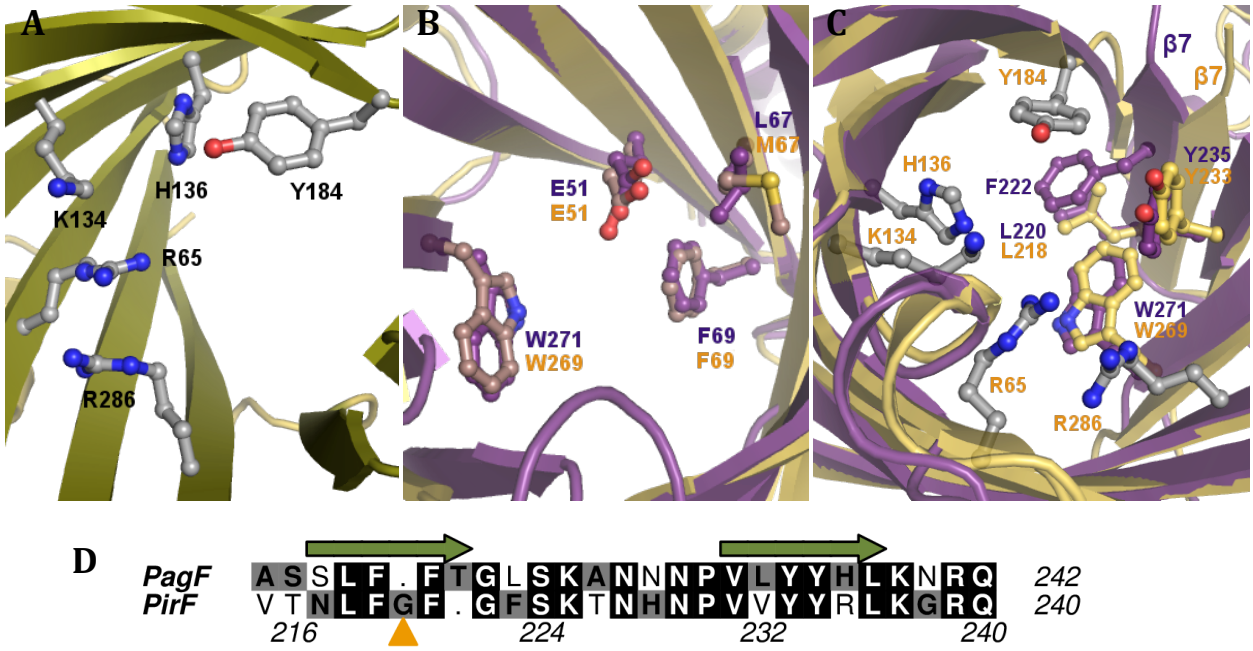

Fig. 4.9 **A.** Residues potentially coordinating the diphosphate moiety of GPP in the substrate-binding channel of PirF. **B.** Superimposition of the residues involved in acceptor Tyr coordination in PagF (purple sticks) and the corresponding residues in PirF (pink sticks). **C.** Superimposition of the substrate-binding channel of PagF (purple) and PirF (gold). Residues for GPP coordination in PirF are shown as gray sticks. Hydrophobic residues at the opposite side are also shown as sticks. The counterpart of the PagF F222 is missing in PirF, causing a substantial outward twist of β7 strand in PirF. **D.** Sequence alignment of the β7 and β8 strands in PagF and PirF. The insertion of Gly at position 220 of PirF forces the side chain of the following F221 to face outside of the barrel. Its counterpart in PagF, F222, faces inside the barrel.

References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Echols, N., Headd, J. J., et al. (2011). The Phenix software for automated determination of macromolecular structures. *Methods , 55* (1), 94-106.

Agarwal, V., Pierce, E., McIntosh, J. A., Schmidt, E. W., & Nair, S. K. (2012). Structures of cyanobactin maturation enzymes define a family of transamidating proteases. *Chemistry & Biology , 19* (11), 1411-1422.

Aminov, R. I. (2010). A brief history of the antibiotic era: lessons learned and challenges for the future. *Frontiers in Microbiology , 1*, 134.

Aminov, R. I. (2009). The role of antibiotics and antibiotic resistance in nature. *Environmental Microbiology , 11* (12), 2970-2988.

Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., et al. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Natural Product Reports , 30* (1), 108-160.

Bachmann, B. O., Van Lanen, S. G., & Baltz, R. H. (2014). Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? *Journal of Industrial Microbiology and Biotechnology , 41* (2), 175-184.

Belshaw, P. J., Roy, R. S., Kelleher, N. L., & Walsh, C. T. (1998). Kinetics and regioselectivity of peptide-to-heterocycle conversions by microcin B17 synthetase. *Chemistry & Biology , 5* (7), 373-384.

Bent, A. F., Koehnke, J., Houssen, W. E., Smith, M. C., Jaspars, M., & Naismith, J. H. (2013). Structure of PatF from Prochloron didemni. *Acta Crystallographica Section F Structural Biology and Crystallization Communications , 69*, 618-623.

Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., et al. (2002). Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature , 417* (6885), 141-147.

Biarrotte-Sorin, S., Maillard, A. P., Delettré, J., Sougakoff, W., Arthur, M., & Mayer, C. (2004). Crystal structures of Weissella viridescens FemX and its complex with UDP-MurNAc-pentapeptide: insights into FemABX family substrates recognition. *Structure , 12* (2), 257-267.

Brachman, P. S. (2003). Infectious diseases-past, present, and future. *International Journal of Epidemiology , 32* (5), 684-686.

Brennan, R. G. (1993). The winged-helix DNA-binding motif: another helix-turn-helix takeoff. *Cell , 74* (5), 773-776.

Breukink, E., van Heusden, H. E., Vollmerhaus, P. J., Swie- zewska, E., Brunner, L., Walker, S., et al. (2003). Lipid II is an intrinsic component of the pore induced by nisin in bacterial membranes. *The Journal of Biological Chemistry , 278* (22), 19898-19903.

Breukink, E., Wiedemann, I., van, K. C., Kuipers, O. P., Sahl, H., & de Kruijff, B. (1999). Use of the cell wall precursor lipid II by a pore-forming peptide antibiotic. *Science , 286* (5448), 2361-2364.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., & Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments

in and around SHARP 2.0. *Acta Crystallographica Section D Biological Crystallography , 59*, 2023-2030.

Brotz, H., Josten, M., Wiedemann, I., Schneider, U., Gotz, F., Bierbaum, G., et al. (1998). Role of lipid-bound peptidoglycan precursors in the formation of pores by nisin, epidermin and other lantibiotics. *Molecular Microbiology , 30* (2), 317-327.

Buchman, G. W., Banerjee, S., & Hansen, J. N. (1988). Structure, expression, and evolution of a gene encoding the precursor of nisin, a small protein antibiotic. *The Journal of Biological Chemistry , 263* (31), 16260-16266.

Burkhart, B. J., Hudson, G. A., Dunbar, K. L., & Mitchell, D. A. (2015). A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. *Nature Chemical Biology , 11* (8), 564-570.

Calteau, A., Fewer, D. P., Latifi, A., Coursin, T., Laurent, T., Jokela, J., et al. (2014). Phylum-wide comparative genomics unravel the diversity of secondary metabolism in Cyanobacteria. *BMC Genomics , 15*, 977.

Casanova, J.-L., & Abel, L. (2013). The genetic theory of infectious disease: a brief history and selected illustrations. *Annual Review of Genomics and Human Genetics , 14*, 215-243.

Castiglione, F., Lazzarini, A., Carrano, L., Corti, E., Ciciliato, I., Gastaldo, L., et al. (2008). Determining the structure and mode of action of microbisporicin, a potent lantibiotic active against multiresistant pathogens. *Chemistry & Biology , 15* (1), 22-31.

Chatterjee, C., Miller, L. M., Leung, Y. L., Xie, L., Yi, M., Kelleher, N. L., et al. (2005). Lacticin 481 synthetase phosphorylates its substrate during lantibiotic production. *Journal of the American Chemical Society , 127* (44), 15332-15333.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., et al. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography , 66*, 12-21.

Claesen, J., & Bibb, M. J. (2011). Biosynthesis and regulation of grisemycin, a new member of the linaridin family of ribosomally synthesized peptides produced by Streptomyces griseus IFO 13350. *Journal of Bacteriology , 193* (10), 2510–2516.

Claesen, J., & Bibb, M. J. (2010). Genome mining and genetic analysis of cypemycin biosynthesis reveal an unusual class of posttranslationally modified peptides. *Proceedings of the National Academy of Sciences of the United States of America , 107* (37), 16297-16302.

Coburn, P. S., & Gilmore, M. S. (2003). The Enterococcus faecalis cytolysin: a novel toxin active against eukaryotic and prokaryotic cells. *Cellular Microbiology , 5* (10), 661-669.

Cotter, P. D., Hill, C., & Ross, R. P. (2005). Bacterial lantibiotics: strategies to improve therapeutic potential. *Current Protein & Peptide Science , 6* (1), 61-75.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D Biological Crystallography , 62*, 1002-1011.

Cui, L., & Su, X. (2009). Discovery, mechanisms of action and combination therapy of artemisinin. *Expert Review of Anti-infective Therapy , 7* (8), 999-1013.

de Ruyter, P. G., Kuipers, O. P., Beerthuyzen, M. M., van Alen-Boerrigter, I., & de Vos, W. M. (1996). Functional analysis of promoters in the nisin gene cluster of Lactococcus lactis. *Journal of Bacteriology , 178* (12), 3434–3439.

Deane, C. D., Melby, J. O., Molohon, K. J., Susarrey, A. R., & Mitchell, D. A. (2013). Engineering unnatural variants of plantazolicin through codon reprogramming. *ACS Chemical Biology , 8* (9), 1998-2008.

Dong, S. H., Tang, W., Lukk, T., Yu, Y., Nair, S. K., & van der Donk, W. A. (2015, July 30). The enterococcal cytolysin synthetase has an unanticipated lipid kinase fold. *Elife* , 4.

Donia, M. S., & Schmidt, E. W. (2011). Linking chemistry and genetics in the growing cyanobactin natural products family. *Chemistry & Biology , 18* (4), 508-519.

Donia, M. S., Hathaway, B. J., Sudek, S., Haygood, M. G., Rosovitz, M. J., Ravel, J., et al. (2006). Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. *Nature Chemical Biology , 2* (12), 729-735.

Donia, M. S., Ravel, J., & Schmidt, E. W. (2008). A global assembly line for cyanobactins. *Nature Chemical Biology , 4* (6), 341-343.

Dorgan, K. M., Wooderchak, W. L., Wynn, D. P., Karschner, E. L., Alfaro, J. F., Cui, Y., et al. (2006). An enzyme-coupled continuous spectrophotometric assay for S-adenosylmethionine-dependent methyltransferases. *Analytical Biochemistry , 350* (2), 249-255.

Doroghazi, J. R., Albright, J. C., Goering, A. W., Ju, K. S., Haines, R. R., Tchalukov, K. A., et al. (2014). A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nature Chemical Biology , 10* (11), 963-968.

Dunbar, K. L., & Mitchell, D. A. (2013). Insights into the mechanism of peptide cyclodehydrations achieved through the chemoenzymatic generation of amide derivatives. *Journal of the American Chemical Society , 135* (23), 8692-8701.

Dunbar, K. L., Chekan, J. R., Cox, C. L., Burkhart, B. J., Nair, S. K., & Mitchell, D. A. (2014). Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis. *Nature Chemical Biology , 10* (10), 823-829.

Dunbar, K. L., Melby, J. O., & Mitchell, D. A. (2012). YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. *Nature Chemical Biology , 8* (6), 569-575.

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography , 60*, 2126-2132.

Engelke, G., Gutowski-Eckel, Z., Kiesau, P., Siegers, K., Hammelmann, M., & Entian, K. D. (1994). Regulation of nisin biosynthesis and immunity in Lactococcus lactis 6F3. *Applied and Environmental Microbiology , 60* (3), 814-825.

Férir, G., Petrova, M. I., Andrei, G., Huskens, D., Hoorelbeke, B., Snoeck, R., et al. (2013). The lantibiotic peptide labyrinthopeptin A1 demonstrates broad anti-HIV and anti-HSV activity with potential for microbicidal applications. *PLoS One , 8* (5), e64010.

Fernandes, P. (2006). Antibacterial discovery and development– the failure of success? *Nature Biotechnology , 24* (12), 1497-1503.

Fischbach, M. A., & Walsh, C. T. (2006). Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery and mechanisms. *Chemical Reviews , 106* (8), 3468-3496.

Fonvielle, M., Chemama, M., Villet, R., Lecerf, M., Bouhss, A., Valéry, J. M., et al. (2009). Aminoacyl-tRNA recognition by the FemXWv transferase for bacterial cell wall synthesis. *Nucleic Acids Research , 37* (5), 1589-1601.

Fonvielle, M., Li de La Sierra-Gallay, I., El-Sagheer, A. H., Lecerf, M., Patin, D., Mellal, D., et al. (2013). The structure of FemX(Wv) in complex with a peptidyl-RNA conjugate: mechanism of aminoacyl transfer from Ala-tRNA (Ala) to peptidoglycan precursors. *Angewandte Chemie International Edition , 52* (28), 7278 –7281.

Foulston, L. C., & Bibb, M. J. (2011). Feed-forward regulation of microbisporicin biosynthesis in Microbispora corallina. *Journal of Bacteriology , 193* (12), 3064-3071.

Foulston, L. C., & Bibb, M. J. (2010). Microbisporicin gene cluster reveals unusual features of lantibiotic biosynthesis in actinomycetes. *Proceedings of the National Academy of Sciences of the United States of America , 107* (30), 13461–13466 .

Francklyn, C. S., & Minajigi, A. (2010). tRNA as an active chemical scaffold for diverse chemical transformations. *FEBS Letters , 584* (2), 366-375.

Garg, N., Salazar-Ocampo, L. M., & van der Donk, W. A. (2013). In vitro activity of the nisin dehydratase NisB. *Proceedings of the National Academy of Sciences of the United States of America , 110* (18), 7258–7263.

Gerlt, J. A., Bouvier, J. T., Davidson, D. B., Imker, H. J., Sadkhin, B., Slater, D. R., et al. (2015). Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochimica et Biophysica Acta , 1854* (8), 1019-1037.

Gonzalez, D. J., Lee, S. W., Hensler, M. E., Markley, A. L., Dahesh, S., Mitchell, D. A., et al. (2010). Clostridiolysin S, a post-translationally modified biotoxin from Clostridium botulinum. *The Journal of Biological Chemistry , 285* (36), 28220-28228.

Goto, Y., Li, B., Claesen, J., Shi, Y., Bibb, M. J., & van der Donk, W. A. (2010). Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights. *PLoS Biology , 8* (3), e1000339.

Goto, Y., Okesli, A., & van der Donk, W. A. (2011). Mechanistic studies of Ser/Thr dehydration catalyzed by a member of the LanL lanthionine synthetase family. *Biochemistry , 50* (5), 891-898.

Gross, E., & Morell, J. L. (1971). The structure of nisin. *Journal of the American Chemical Society , 93* (18), 4634-4635.

Hannoush, R. N., & Sun, J. (2010). The chemical toolbox for monitoring protein fatty acylation and prenylation. *Nature Chemical Biology , 6* (7), 498-506.

Harvey, A. L., Edrada-Ebel, R., & Quinn, R. J. (2015). The re-emergence of natural products for drug discovery in the genomics era. *Nature Reviews Drug Discovery , 14*, 111-129.

Hasper, H. E., de Kruijff, B., & Breukink, E. (2004). Assembly and stability of nisin-lipid II pores. *Biochemistry , 43* (36), 11567-11575.

Hasper, H. E., Kramer, N. E., Smith, J. L., Hillman, J. D., Zachariah, C., Kuipers, O. P., et al. (2006). An alternative bactericidal mechanism of action for lantibiotic peptides that target lipid II. *Science , 313* (5793), 1636-1637.

Heide, L. (2009). Prenyl transfer to aromatic substrates: genetics and enzymology. *Current Opinion in Chemical Biology , 13* (2), 171–179.

Horn, M. J., Jones, D. B., & Ringel, S. J. (1941). Isolation of a new sulfur-containing amino acid (lanthionine) from sodium carbonate-treated wool. *Journal of Biological Chemistry , 138*, 141-149.

Houssen, W. E., & Jaspars, M. (2010). Azole-based cyclic peptides from the sea squirt Lissoclinum patella: old scaffolds, new avenues. *Chembiochem , 11* (13), 1803–1815.

Houssen, W. E., Bent, A. F., McEwan, A. R., Pieiller, N., Tabudravu, J., Koehnke, J., et al. (2014). An efficient method for the in vitro production of azol(in)e-based cyclic peptides. *Angewandte Chemie International Edition , 53* (51), 14171-41417.

Houssen, W. E., Koehnke, J., Zollman, D., Vendome, J., Raab, A., Smith, M. C., et al. (2012). The discovery of new cyanobactins from Cyanothece PCC 7425 defines a new signature for processing of patellamides. *Chembiochem , 13* (18), 2683-2689.

Hsu, S. T., Breukink, E., Tischenko, E., Lutters, M. A., de Kruijff, B., Kaptein, R., et al. (2004). The nisin-lipid II complex reveals a pyrophosphate cage that provides a blueprint for novel antibiotics. *Nature Structural & Molecular Biology , 11* (10), 963-967.

Ireland, C., & Scheuer, P. J. (1980). Ulicyclamide and ulithiacyclamide, two new small peptides from a marine tunicate. *Journal of the American Chemical Society , 102* (17), 5688-5691.

Jabés, D., Brunati, C., Candiani, G., Riva, S., Romanó, G., & Donadio, S. (2011). Efficacy of the new lantibiotic NAI-107 in experimental infections induced by multidrug-resistant Gram-positive pathogens. *Antimicrobial Agents and Chemotherapy , 55* (4), 1671-1676.

Kabsch, W. (2010). XDS. *Acta Crystallographica Section D Biological Crystallography , 66*, 125-132.

Kaletta, C., & Entian, K. D. (1989). Nisin, a peptide antibiotic: Cloning and sequencing of the nisA gene and posttranslational processing of its peptide product. *Journal of Bacteriology , 171* (3), 1597-1601.

Kalyon, B., Helaly, S. E., Scholz, R., Nachtigall, J., Vater, J., Borriss, R., et al. (2011). Plantazolicin A and B: structure elucidation of ribosomally synthesized thiazole/oxazole peptides from Bacillus amyloliquefaciens FZB42. *Organic Letters , 13* (12), 2996-2999.

Karamanou, M., Panayiotakopoulos, G., Tsoucalas, G., Kousoulis, A. A., & Androutsos, G. (2012). From miasmas to germs: a historical approach to theories of infectious disease transmission. *Le Infezioni in Medicina , 20* (1), 58-62.

Kelleher, N. L., Hendrickson, C. L., & Walsh, C. T. (1999). Posttranslational heterocyclization of cysteine and serine residues in the antibiotic microcin B17: distributivity and directionality. *Biochemistry , 38* (47), 15623-15630.

Kellner, R., Jung, G., Hörner, T., Zähner, H., Schnell, N., Entian, K. D., et al. (1988). Gallidermin : a new lanthionine-containing polypeptide antibiotic. *European Journal of Biochemistry , 177* (1), 53-59.

Khaldi, N., Seifuddin, F. T., Turner, G., Haft, D., Nierman, W. C., Wolfe, K. H., et al. (2010). SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genetics and Biology , 47* (9), 736–741.

Khusainov, R., Moll, G. N., & Kuipers, O. P. (2013). Identification of distinct nisin leader peptide regions that determine interactions with the modification enzymes NisB and NisC. *FEBS Open Bio , 3*, 237–242.

Kleywegt, G. J., & Brünger, A. T. (1996). Checking your imagination: applications of the free R value. *Structure , 4* (8), 897-904.

Kluskens, L. D., Kuipers, A., Rink, R., de Boef, E., Fekken, S., Driessen, A. J., et al. (2005). Post-translational modification of therapeutic peptides by NisB, the dehydratase of the lantibiotic nisin. *Biochemistry , 44* (38), 12827-12834.

Knerr, P. J., & van der Donk, W. A. (2012). Discovery, biosynthesis and engineering of lanthipeptides. *Annual Review of Biochemistry , 81*, 479-505.

Koehnke, J., Bent, A. F., Zollman, D., Smith, K., Houssen, W. E., Zhu, X., et al. (2013). The cyanobactin heterocyclase enzyme: a processive adenylase that operates with a defined order of reaction. *Angewandte Chemie International Edition , 52* (52), 13991-13996.

Koehnke, J., Bent, A., Houssen, W. E., Zollman, D., Morawitz, F., Shirran, S., et al. (2012). The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain. *Nature Structural & Molecular Biology , 19* (8).

Koehnke, J., Mann, G., Bent, A. F., Ludewig, H., Shirran, S., Botting, C., et al. (2015). Structural analysis of leader peptide binding enables leader-free cyanobactin processing. *Nature Chemical Biology , 11* (8), 558-563.

Koponen, O., Tolonen, M., Qiao, M., Wahlstrom, G., Helin, J., & Saris, P. E. (2002). NisB is required for the dehydration and NisC for the lanthionine formation in the post-translational modification of nisin. *Microbiology , 148* (11), 3561-3568.

Kozbial, P. Z., & Mushegian, A. R. (2005). Natural history of S-adenosylmethionine-binding proteins. *BMC Structural Biology , 5*, 19.

Krawczyk, B., Ensle, P., Müller, W. M., & Süssmuth, R. D. (2012). Deuterium labeled peptides give insights into the directionality of class III lantibiotic synthetase LabKC. *Journal of the American Chemical Society , 134* (24), 9922–9925.

Krawczyk, B., Völler, G. H., Völler, J., Ensle, P., & Süssmuth, R. D. (2012). Curvopeptin: a new lanthionine-containing class III lantibiotic and its co-substrate promiscuous synthetase. *Chembiochem , 13* (14), 2065-2071.

Kuipers, O. P., Beerthuyzen, M. M., de Ruyter, P. G., Luesink, E. J., & de Vos, W. M. (1995). Autoregulation of nisin biosynthesis in Lactococcus lactis by signal transduction. *The Journal of Biological Chemistry , 270* (45), 27299-32704.

Kuipers, O. P., Beerthuyzen, M. M., Siezen, R. J., & de Vos, W. (1993). Characterization of the nisin gene cluster nisABTCIPR of Lactococcus lactis. Requirement of expres- sion of the nisA and nisI genes for development of immunity. *European Journal of Biochemistry , 216* (1), 281-291.

Kuzuyama, T., Noel, J. P., & Richard, S. B. (2005). Structural basis for the promiscuous biosynthetic prenylation of aromatic natural products. *Nature , 435* (7044), 983-987.

Laskowski, R. A., & Swindells, M. B. (2011). LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *Journal of Chemical Informaion and Modeling , 51* (10), 2778-2786.

Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR , 8* (4), 477-486.

Lee, J., Hao, Y., Blair, P. M., Melby, J. O., Agarwal, V., Burkhart, B. J., et al. (2013). Structural and functional insight into an unexpectedly selective N-methyltransferase involved in plantazolicin biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America , 110* (32), 12954–12959.

Lee, J., McIntosh, J., Hathaway, B. J., & Schmidt, E. W. (2009). Using marine natural products to discover a protease that catalyzes peptide macrocyclization of diverse substrates. *Journal of the American Chemical Society , 131* (6), 2122-2124.

Lee, M., Ihnken, L., You, Y., McClerren, A., van der Donk, W., & Kelleher, N. (2009). Distributive and directional behavior of lantibiotic synthetases revealed by high-resolution tandem mass spectrometry. *Journal of the American Chemical Society , 131* (34), 12258-12264.

Lee, S., Mitchell, D., Markley, A., Hensler, M., Gonzalez, D., Wohlrab, A., et al. (2008). Discovery of a widely distributed toxin biosynthetic gene cluster. *Proceedings of the National Academy of Sciences of the United States of America , 105* (15), 5879-5884.

Leikoski, N., Fewer, D. P., Jokela, J., Alakoski, P., Wahlsten, M., & Sivonen, K. (2012). Analysis of an inactive cyanobactin biosynthetic gene cluster leads to discovery of new natural products from strains of the genus Microcystis. *PLoS One , 7* (8), e43002.

Leikoski, N., Liu, L., Jokela, J., Wahlsten, M., Gugger, M., Calteau, A., et al. (2013). Genome mining expands the chemical diversity of the cyanobactin family to include highly modified linear peptides. *Chemistry & Biology , 20* (8), 1033-1043.

Letzel, A.-C., Pidot, S. J., & Hertweck, C. (2014). Genome mining for ribosomally synthesized and post-translationally modified peptides (RiPPs) in anaerobic bacteria. *BMC Genomics , 15*, 983.

Li, B., Sher, D., Kelly, L., Shi, Y., Huang, K., Knerr, P. J., et al. (2010). Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America , 107* (23), 10430–10435.

Li, B., Yu, J. P., Brunzelle, J. S., Moll, G. N., van der Donk, W. A., & Nair, S. K. (2006). Structure and mechanism of the lantibiotic cyclase involved in nisin biosynthesis. *Science , 311* (5766), 1464-1467.

Li, C., & Kelly, W. L. (2010). Recent advances in thiopeptide antibiotic biosynthesis. *Natural Product Reports , 27* (2), 153-164.

Li, C., Zhang, F., & Kelly, W. L. (2011). Heterologous production of thiostrepton A and biosynthetic engineering of thiostrepton analogs. *Molecular BioSystems , 7* (1), 82-90.

Li, J. W.-H., & Vederas, J. C. (2009). Drug discovery and natural products: end of an era or an endless frontier? *Science , 325* (5937), 161-165.

Li, Y., Milne, J., Madison, L., Kolter, R., & Walsh, C. (1996). From peptide precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase. *Science , 274* (5290), 1188-1193.

Liang, P. H., Ko, T. P., & Wang, A. H. (2002). Structure, mechanism and function of prenyltransferases. *European Journal of Biochemistry , 269* (14), 3339-3354.

Lubelski, J., Khusainov, R., & Kuipers, O. P. (2009). Directionality and coordination of dehydration and ring formation during biosynthesis of the lantibiotic nisin. *The Journal of Biological Chemistry , 284* (38), 25962-25972.

Müller, W. M., Ensle, P., Krawczyk, B., & Süssmuth, R. D. (2011). Leader peptide-directed processing of labyrinthopeptin A2 precursor peptide by the modifying enzyme LabKC. *Biochemistry , 50* (39), 8362–8373.

Münch, D., Müller, A., Schneider, T., Kohl, B., Wenzel, M., Bandow, J. E., et al. (2014). The lantibiotic NAI-107 binds to bactoprenol-bound cell wall precursors and impairs membrane functions. *The Journal of Biological Chemistry , 289* (17), 12063-12076.

Magnuson, R., Solomon, J., & Grossman, A. D. (1994). Biochemical and genetic characterization of a competence pheromone from B. subtilis. *Cell , 77* (2), 207-216.

Mainardi, J. L., Villet, R., Bugg, T. D., Mayer, C., & Arthur, M. (2008). Evolution of peptidoglycan biosynthesis under the selective pressure of antibiotics in Gram-positive bacteria. *FEMS Microbiology Reviews , 32* (2), 386-408.

Marques, J. C., Lamosa, P., Russell, C., Ventura, R., Maycock, C., Semmelhack, M. F., et al. (2011). Processing the interspecies quorum-sensing signal autoinducer-2 (AI-2): characterization of phospho-(S)-4,5-dihydroxy-2,3-pentanedione isomerization by LsrG protein. *The Journal of Biological Chemistry , 286* (20), 18331-18343.

Mattick, A. T., & Hirsch, A. (1944). A powerful inhibitory substance produced by group N Streptococci. *Nature* (3913), 551-551.

Mavaro, A., Abts, A., Bakkes, P. J., Moll, G. N., Driessen, A. J., Smits, S. H., et al. (2011). Substrate recognition and specificity of the NisB protein, the lantibiotic dehydratase involved in nisin biosynthesis. *The Journal of Biological Chemistry , 286* (35), 30552–30560.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography , 40*, 658-674.

McIntosh, J. A., & Schmidt, E. W. (2010). Marine molecular machines: heterocyclization in cyanobactin biosynthesis. *Chembiochem , 11* (10), 1413-1421.

McIntosh, J. A., Donia, M. S., & Schmidt, E. W. (2010). Insights into heterocyclization from two highly similar enzymes. *Journal of the American Chemical Society , 132* (12), 4089-4091.

McIntosh, J. A., Donia, M. S., Nair, S. K., & Schmidt, E. W. (2011). Enzymatic basis of ribosomal peptide prenylation in cyanobacteria. *Journal of the American Chemical Society , 133* (34), 13698-13705.

McIntosh, J. A., Robertson, C. R., Agarwal, V., Nair, S. K., Bulaj, G. W., & Schmidt, E. W. (2010). Circular logic: nonribosomal peptide-like macrocyclization with a ribosomal peptide catalyst. *Journal of the American Chemical Society , 132* (44), 15499-15501.

Meindl, K., Schmiederer, T., Schneider, K., Reicke, A., Butz, D., Keller, S., et al. (2010). Labyrinthopeptins: a new class of carbacyclic lantibiotics. *Angewandte Chemie International Edition , 49* (6), 1151–1154.

Melby, J. O., Dunbar, K. L., Trinh, N. Q., & Mitchell, D. A. (2012). Selectivity, directionality, and promiscuity in peptide processing from a Bacillus sp. Al Hakam cyclodehydratase. *Journal of the American Chemical Society , 134* (11), 5309-5316.

Melby, J. O., Li, X., & Mitchell, D. A. (2014). Orchestration of enzymatic processing by thiazole/oxazole-modified microcin dehydrogenases. *Biochemistry , 53* (2), 413-422.

Melby, J. O., Nard, N. J., & Mitchell, D. A. (2011). Thiazole/oxazole-modified microcins: complex natural products from ribosomal templates. *Current Opinion in Chemical Biology , 15* (3), 369-378.

Metzger, U., Keller, S., Stevenson, C. E., Heide, L., & Lawson, D. M. (2010). Structure and mechanism of the magnesium-independent aromatic prenyltransferase CloQ from the clorobiocin biosynthetic pathway. *Journal of Molecular Biology , 404* (4), 611-626.

Metzger, U., Schall, C., Zocher, G., Unsöld, I., Stec, E., Li, S. M., et al. (2009). The structure of dimethylallyl tryptophan synthase reveals a common architecture of aromatic prenyltransferases in fungi and bacteria. *Proceedings of the National Academy of Sciences of the United States of America , 106* (34), 14309-14314.

Milne, J. C., Roy, R. S., Eliot, A. C., Kelleher, N. L., Wokhlu, A., Nickels, B., et al. (1999). Cofactor requirements and reconstitution of microcin B17 synthetase: a multienzyme complex that catalyzes the formation of oxazoles and thiazoles in the antibiotic microcin B17. *Biochemistry , 38* (15), 4768-4781.

Mitchell, D. A., Lee, S. W., Pence, M. A., Markley, A. L., Limm, J. D., Nizet, V., et al. (2009). Structural and functional dissection of the heterocyclic peptide cytotoxin streptolysin S. *The Journal of Biological Chemistry , 284* (19), 13004-13012.

Mohr, K. I., Volz, C., Jansen, R., Wray, V., Hoffmann, J., Bernecker, S., et al. (2015, July 24). Pinensins: the first antifungal lantibiotics. *Angewandte Chemie International Edition* .

Molohon, K. J., Melby, J. O., Lee, J., Evans, B. S., Dunbar, K. L., Bumpus, S. B., et al. (2011). Structure determination and interception of biosynthetic intermediates for the plantazolicin class of highly discriminating antibiotics. *ACS Chemical Biology , 16* (6), 1307-1313.

Mukherjee, S., & van der Donk, W. A. (2014). Mechanistic studies on the substrate-tolerant lanthipeptide synthetase ProcM. *Journal of the American Chemical Society , 136* (29), 10450-10459.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., et al. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D Biological Crystallography , 67*, 355-367.

Nolan, E. M., & Walsh, C. T. (2009). How nature morphs peptide scaffolds into antibiotics. *Chembiochem , 10* (1), 34-53.

Ortega, M. A., Hao, Y., Zhang, Q., Walker, M. C., & van der Donk, W. A. (2015). Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. *Nature , 517* (7535), 509-512.

Otwinowski, Z., & Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. In C. W. Carter, & R. M. Sweet (Eds.), *Methods in Enzymology, part A* (Vol. 276, pp. 307-326). New York: Academic Press.

Owen, J. G., Charlop-Powers, Z., Smith, A. G., Ternei, M. A., Calle, P. Y., Reddy, B. V., et al. (2015). Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. *Proceedings of the National Academy of Sciences of the United States of America , 112* (14), 4221-4226.

Palsuledesai, C. C., & Distefano, M. D. (2015). Protein prenylation: enzymes, therapeutics, and biotechnology applications. *ACS Chemical Biology , 10* (1), 51-62.

Pan, S. J., & Link, A. J. (2011). Sequence diversity in the lasso peptide framework: discovery of functional Microcin J25 variants with multiple amino acid substitutions. *Journal of the American Chemical Society , 133* (13), 5016–5023.

Pascolutti, M., & Quinn, R. J. (2014). Natural products as lead structures: chemical transformations to create lead-like libraries. *Drug Discovery Today , 19* (3), 215-221.

Piwowarska, N. A., Banala, S., Overkleeft, H. S., & Süssmuth, R. D. (2013). Arg-Thz is a minimal substrate for the Na,Na-arginyl methyltransferase involved in the biosynthesis of plantazolicin. *Chemical Communications , 49* (91), 10703-10705.

Plat, A., Kluskens, L. D., Kuipers, A., Rink, R., & Moll, G. N. (2011). Requirements of the engineered leader peptide of nisin for inducing modification, export, and cleavage. *Applied and Environmental Microbiology , 77* (2), 604-611.

Qiao, M., & Saris, P. E. (1996). Evidence for a role of NisT in transport of the lantibiotic nisin produced by Lactococcus lactis N8. *FEMS Microbiology Letters , 144* (1), 89-93.

Qiao, M., Immonen, T., Koponen, O., & Saris, P. E. (1995). The cellular location and effect on nisin immunity of the NisI protein from Lactococcus lactis N8 expressed in Escherichia coli and L. lactis. *FEMS Microbiology Letters , 131* (1), 75-80.

Qiao, M., Ye, S., Koponen, O., Ra, R., Usabiaga, M., Immonen, T., et al. (1996). Regulation of the nisin operons in Lactococcus lactis N8. *Journal of Applied Bacteriology , 80* (6), 626-634.

Reddy, B. V., Kallifidas, D., Kim, J., Charlop-Powers, Z., Feng, Z., & Brady, S. F. (2012). Natural product biosynthetic gene diversity in geographically distinct soil microbiomes. *Applied and Environmental Microbiology , 78* (10), 3744-3752.

Rink, R. ,., de Boef, E., Leenhouts, K. J., Driessen, A. J., Moll, G. N., & Kuipers, O. P. (2005). Lantibiotic structures as guidelines for the design of peptides that can be modified by lantibiotic enzymes. *Biochemistry , 44* (24), 8873-8882.

Rink, R., Wierenga, J., Kuipers, A., Kluskens, L. D., Dries-sen, A. J., Kuipers, O. P., et al. (2007). Production of dehydroamino acid-containing peptides by Lactococcus lactis. *Applied and Environmental Microbiology , 73* (6), 1792-1796.

Rogers, L. A. (1928). The inhibiting effect of Streptococcus lactis on Lactobacillus bulgaricus. *Journal of Bacteriology , 16* (5), 321-325.

Roy, R. S., Kim, S., Baleja, J. D., & Walsh, C. T. (1998). Role of the microcin B17 propeptide in substrate recognitibn: solution structure and mutational analysis of McbA1-26. *Chemistry & Biology , 5* (4), 217-228.

Ruffner, D. E., Schmidt, E. W., & Heemstra, J. R. (2015). Assessing the combinatorial potential of the RiPP cyanobactin tru pathway. *ACS Synthetic Biology , 4* (4), 482-492.

Sardar, D., Lin, Z., & Schmidt, E. W. (2015). Modularity of RiPP enzymes enables designed synthesis of decorated peptides. *Chemistry & Biology , 22* (7), 907-916.

Sardar, D., Pierce, E., McIntosh, J. A., & Schmidt, E. W. (2015). Recognition sequences and substrate evolution in cyanobactin biosynthesis. *ACS Synthetic Biology , 4* (2), 167-176.

Schmidt, E. W., Nelson, J. T., Rasko, D. A., Sudek, S., Eisen, J. A., Haygood, M. G., et al. (2005). Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. *Proceedings of the National Academy of Sciences of the United States of America , 102* (20), 7315-7320.

Scholz, R., Molohon, K. J., Nachtigall, J., Vater, J., Markley, A. L., Süssmuth, R. D., et al. (2011). Plantazolicin, a novel microcin B17/streptolysin S-like natural product from Bacillus amyloliquefaciens FZB42. *Journal of Bacteriology , 193* (1), 215-224.

Severina, E., Severin, A., & Tomasz, A. (1998). Antibacterial efficacy of nisin against multidrug-resistant Gram-positive pathogens. *Journal of Antimicrobial Chemotherapy , 41* (3), 341-347.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research , 13* (11), 2498-2504.

Sharma, A., Blair, P. M., & Mitchell, D. A. (2013). Synthesis of plantazolicin analogues enables dissection of ligand binding interactions of a highly selective methyltransferase. *Organic Letters , 15* (19), 5076–5079.

Shazman, S., & Mandel-Gutfreund, Y. (2008). Classifying RNA-binding proteins based on electrostatic properties. *PLoS Computational Biology , 4* (8), e1000146.

Shulman, S. T. (2004). The history of pediatric infectious diseases. *Pediatric Research , 55* (1), 163-176.

Siegers, K., & Entian, K. D. (1995). Genes involved in immunity to the lantibiotic nisin produced by Lactococcus lactis 6F3. *Applied and Environmental Microbiology , 61* (3), 1082-1089.

Siegers, K., Heinzmann, S., & Entian, K. D. (1996). Biosynthesis of lantibiotic nisin. Posttranslational modification of its prepeptide occurs at a multimeric membrane-associated lanthionine synthetase complex. *The Journal of Biological Chemistry , 271* (21), 12294-12301.

Singh, M., & Sareen, D. (2014). Novel LanT associated lantibiotic clusters identified by genome database mining. *PLoS One , 9* (3), e91352.

Sivonen, K., Leikoski, N., Fewer, D. P., & Jokela, J. (2010). Cyanobactins—ribosomal cyclic peptides produced by cyanobacteria. *Applied Microbiology and Biotechnology , 86* (5), 1213-1225.

Solecki, R. S. (1975). Shanidar IV, a Neanderthal flower burial in northern Iraq. *Science , 190* (4217), 880-881.

Stein, T., Heinzmann, S., Solovieva, I., & Entian, K. D. (2003). Function of Lactococcus lactis nisin immunity genes nisI and nisFEG after coordinated expression in the surrogate host Bacillus subtilis. *The Journal of Biological Chemistry , 278* (1), 89-94.

Stepper, J., Shastri, S., Loo, T. S., Preston, J., Novak, P., Man, P., et al. (2011). Cysteine S-glycosylation, a new post-translational modification found in glycopeptide bacteriocins. *FEBS Letters , 585* (4), 645-650.

Stern, A. M., & Markel, H. (2005). The history of vaccines and immunization: familiar patterns, new challenges. *Health Affairs , 24* (3), 611-621.

Sukuru, S. C., Jenkins, J. L., Beckwith, R. E., Scheiber, J., Bender, A., Mikhailov, D., et al. (2009). Plate-based diversity selection based on empirical HTS data to enhance the number of hits and their chemical diversity. *Journal of Biomolecular Screening , 14* (6), 690-699.

Takala, T. M., Koponen, O., Qiao, M., & Saris, P. E. (2004). Lipid-free NisI: Interaction with nisin and contribution to nisin immunity via secretion. *FEMS Microbiology Letters , 237* (1), 171-177.

Takano, E. (2006). Gamma-butyrolactones: Streptomyces signalling molecules regulating antibiotic production and differentiation. *Current Opinion in Microbiology , 9* (3), 287-294.

Tello, M., Kuzuyama, T., Heide, L., Noel, J. P., & Richard, S. B. (2008). The ABBA family of aromatic prenyltransferases: broadening natural product diversity. *Cellular and Molecular Life Sciences , 65* (10), 1459-1463.

Thorn, A., & Sheldrick, G. M. (2013). Extending molecular-replacement solutions with SHELXE. *Acta Crystallographica Section D Biological Crystallography , 69*, 2251-2256.

Tianero, M. D., Donia, M. S., Young, T. S., Schultz, P. G., & Schmidt, E. W. (2012). Ribosomal route to small-molecule diversity. *Journal of the American Chemical Society , 134* (1), 418-425.

Todd, E. W. (1938). The differentiation of two distinct serologic varieties of streptolysin, streptolysin O and streptolysin S. *The Journal of Pathology and Bacteriology , 47* (3), 423-445.

Tsuji, F., Ishihara, A., Nakagawa, A., Okada, M., Kitamura, S., Kanamaru, K., et al. (2012). Lack of the consensus sequence necessary for tryptophan prenylation in the ComX pheromone precursor. *Bioscience, Biotechnology, and Biochemistry , 76* (8), 1492-1496.

U.S. Centers for Disease Control and Prevention (CDC). (2013). *Antibiotic resistance threats in the United States, 2013.*

Völler, G. H., Krawczyk, J. M., Pesic, A., Krawczyk, B., Nachtigall, J., & Süssmuth, R. D. (2012). Characterization of new class III lantibiotics--erythreapeptin, avermipeptin and griseopeptin from Saccharopolyspora erythraea, Streptomyces avermitilis and Streptomyces griseus demonstrates stepwise N-terminal leader processing. *Chembiochem , 13* (8), 1174-1183.

van der Meer, J. R., Polman, J., Beerthuyzen, M. M., Siezen, R. J., Kuipers, O. P., & de Vos, W. M. (1993). Characterization of the Lactococcus lactis nisin A operon genes nisP, encoding a

subtilisin-like serine protease involved in precursor processing, and nisR, encoding a regulatory protein involved in nisin biosynthesis. *Journal of Bacteriology , 175* (9), 2578-2588.

van Heusden, H. E., de Kruijff, B., & Breukink, E. (2002). Lipid II induces a transmembrane orientation of the pore- forming peptide lantibiotic nisin. *Biochemistry , 41* (40), 12171-12178.

Vizán, J. L., Hernández-Chico, C., del Castillo, I., & Moreno, F. (1991). The peptide antibiotic microcin B17 induces double-strand cleavage of DNA mediated by E. coli DNA gyrase. *The EMBO Journal , 10* (2), 467-476.

Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., et al. (2011). Data processing and analysis with the autoPROC toolbox. *Acta Crystallographica Section D Biological Crystallography , 67*, 293-302.

Walter, T. S., Meier, C., Assenberg, R., Au, K. F., Ren, J., Verma, A., et al. (2006). Lysine methylation as a routine rescue strategy for protein crystallization. *Structure , 14* (11), 1617-1622.

Wang, H., & van der Donk, W. A. (2012). Biosynthesis of the class III lantipeptide catenulipeptin. *ACS Chemical Biology , 7* (9), 1529–1535.

Watts, S. (1997). *Epidemics and History: Disease, Power and Imperialism.* New Haven, Connecticut: Yale University Press.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., et al. (2015, May 6). antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research* .

Xie, L., Miller, L. M., Chatterjee, C., Averin, O., Kelleher, N. L., & van der Donk, W. A. (2004). Lacticin 481: in vitro reconstitution of lantibiotic synthetase activity. *Science , 303* (5658), 679-681.

Zhang, G., Li, Y., Fang, L., & Pfeifer, B. A. (2015). Tailoring pathway modularity in the biosynthesis of erythromycin analogs heterologously engineered in E. coli. *Science Advances , 1* (4), e1500077.

Zhang, K. Y., Cowtan, K., & Main, P. (1997). Combining constraints for electron-density modification. *Methods in Enzymology , 277*, 53-64.

Zhang, Q., & van der Donk, W. A. (2012). Catalytic promiscuity of a bacterial a-N-methyltransferase. *FEBS Letters , 586* (19), 3391–3397.

Zverina, E. A., Lamphear, C. L., Wright, E. N., & Fierke, C. A. (2012). Recent advances in protein prenyltransferases: substrate identification, regulation, and disease interventions. *Current Opinion in Chemical Biology , 12* (5-6), 544-552.